5-22-2019

Exploring the Usage of Multiple Learning Systems in Learning to Read

TONG LI
University of Connecticut - Storrs, litongpsy@gmail.com

Exploring the Usage of Multiple Learning Systems in Learning to Read

Tong Li, PhD

University of Connecticut, 2019

How people learn to read is an interesting question which has been investigated by many studies with various approaches. Some recent studies have related learning to read with domain-general abilities and have found a positive relationship between statistical learning and learning to read, as well as between procedural learning and learning to read. However, evidence on these relationships is still inconsistent, which probably because reading, statistical learning and procedural learning are componential capabilities. The current study provided another approach to explore how people learn to read, especially how to learn the orthography-phonology (O-P) and orthography-semantics (O-S) correspondences, with multiple learning systems: reflective learning which mostly underlies rule-based learning, and reflexive learning which mostly underlies information-integration. An artificial orthography learning paradigm (AOL) was used as the measure of learning to read with statistical regularities built in O-P and O-S correspondences. In Experiment 1, different manipulations were used on AOL tasks to disrupt either reflective learning or reflexive learning. Disrupting reflective learning significantly impaired performance on AOL tasks, and the O-S learning was more impaired than O-P learning. However, disrupting reflexive learning did not affect overall learning. Experiment 2 further examined the relationship between reflective/reflexive learning and the individual differences in learning to read, this time reflective and reflexive learning were directly measured. Reflective learning was a significant and robust predictor for AOL performance, but reflexive learning was only a predictor to AOL training but not categorization. A trend of competition between the two learning types was also shown by the interaction between them. In addition, reflexive learning but not reflective learning predicted visual statistical leaning, and working memory was found to be positively correlated with both types of learning. Taken together, this study showed that reflective learning was engaged in learning to read with the AOL tasks. The engagement of reflexive learning was also possible, but probably was diminished by the competition between the two learnings and the paradigm of tasks. Although we should be cautious when generalize the findings to a broader question of learning to read, this study provides insights in understanding reading acquisition and education.

Keywords: reflective learning, reflexive learning, artificial orthography learning, individual differences

Exploring the Usage of Multiple Learning Systems in Learning to Read

Tong Li

B.S., Peking University, 2007

M.Phil., The Chinese University of Hong Kong, 2012

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at the

University of Connecticut

2019

APPROVAL PAGE

Doctor of Philosophy Dissertation

Exploring the Usage of Multiple Learning Systems in Learning to Read

Presented by

Tong Li, B.S., M.Phil.

Major Advisor

Jay G. Rueckl

Associate Advisor

Kenneth Pugh

Associate Advisor

James Magnuson

Associate Advisor

Nicole Landi

University of Connecticut
2019

ACKNOWLEDGEMENTS

I would like to give my sincere gratitude and appreciation to my advisor, Dr. Jay Rueckl, for his valuable insight, guidance, and support for my research projects for years.

And I want to extend my gratitude to my committee members, Dr. Kenneth Pugh, Dr. James Magnuson and Dr. Nicole Landi. This research would not have been possible without their advice.

I also want to give my thanks to Dr. Julie Brown, who helped me copyedit my draft and who is always a wonderful audience. And I would also like to thank other faculty members and fellow graduate students in my program for their support in these years.

To my parents, thank you for your understanding and support. I feel sorry I cannot be with you for so many years.

Finally, to my girlfriend Suey Han, thank you for your support in my hardest time. Love you.

Table of Content

List of Tables

List of Figures

Introduction

Learning regularities as a core problem in learning to read

Word reading requires multiple cognitive skills, and a deficit in any of these skills could make reading more difficult (Castles & Coltheart, 1993; Coltheart, 2015; Pennington, 2006; Ziegler, Castel, Pech-Georgel, George, Alario, & Perry, 2008). Many studies have found that reading performance and reading acquisition are affected by several literacy-related abilities. For instance, phonological awareness, letter-sound knowledge, and rapid automatized naming have been found to be robust predictors to children's reading performance and development (e.g., Araújo, Reis, Petersson & Faísca, 2015; Hulme, Bowyer-Crane, Carroll, Duff & Snowling, 2012; Hulme & Snowling, 2013; Melby-Lervåg, Lyster & Hulme, 2012; Norton & Wolf, 2012), although their relative importance as predictors of reading ability across different orthographies may vary (e.g., Caravolas et al., 2012; Kirby, Georgiou, Martinussen & Parrila, 2012; Wimmer, Mayringer & Landerl, 2000).

Apart from the research investigating the relationship between literacy-related skills and reading, some other studies have focused on the underlying processes of reading. Researchers have noticed that rich statistical structure exists in both oral and written language, and behavioral studies have shown that these regularities affect reading performance. For example, 'ill' is always pronounced as /ɪl/ at the end of a word, like in *mill*, *pill*, and *till*, which reflects a highly regular correspondence in orthography-to-phonology (O-P) mapping. The words with a reliable O-P mapping are usually read faster and more accurately than words with an inconsistent O-P mapping (e.g., *pint;* Glushko, 1979; Jared, 2002). The orthography-to-semantics (O-S) mapping can also have statistical structures that affect reading performance. For example, the *-er* in *worker* and *teacher* means a person who takes the action indicated by the word stem. In lexical decision tasks which require individuals to identify pseudo words, it is more difficult to decide that the pseudowords containing real word stems are not words (e.g., Taft & Forster, 1995). Usually the O-P mapping has more regularities compared to O-S mapping, which may result in different reading performance through O-P and O-S pathways. As suggested by the Dual-Route Cascade model (DRC, e.g., Coltheart, Rastle, Perry, Langdon & Ziegler, 2001), reading through the O-P mapping can be achieved by two different systems: the grapheme-phoneme rule system that is used to read regular words and nonwords and the holistic lexicon system that is used to read irregular words. Reading through the O-S mapping cannot use the grapheme-phoneme rule system but it can only rely on a holistic system. However, according to the triangle model (e.g., Harm & Seidenberg, 2004), reading is shaped by the statistical structure of the writing system, and reading through the O-P pathway and the O-S pathway should use the same system.

In support of the triangle model behavioral studies have shown that the statistical structure between orthography, phonology, and semantics is used by learners when learning to read. Developmental studies have found that young children's spellings are influenced by both legal combinations of letters and by morphological patterns (e.g., Deacon, Conrad & Pacton, 2008; Hayes, Treiman & Kessler, 2006). As they are exposed to more words, children gradually exhibit the ability to decide on spellings using statistical regularities in context, e.g., judge the spelling of the vowel based on the coda of the nonword (Treiman & Kessler, 2006), and children in 1st and 2nd grade already show a preference for more frequently used derived allomorphs compared to the ones not so frequently seen (e.g., using -er rather than -or) when choosing the correct ending of words (Deacon & Leung, 2013). Using artificial language learning tasks, researchers have found that adults learn more effectively when the languages have statistical structures between orthography, phonology and semantics (e.g., Deng et al., 2011; Havas et al., 2015; Merkx et al., 2011; Rueckl & Dror, 1994; Taylor et al., 2011; Trudeau, 2006). Computational modelling which successfully simulates human reading and reading acquisition behavior in both normal developing and dyslexic individuals also suggests that reading performance and learning to read require using regularities which underlie correspondences between orthography, phonology, and semantics, although models may differ in what routes are used (e.g., Coltheart, Curtis, Atkins & Haller, 1993; Coltheart, Rastle, Perry, Langdon & Ziegler, 2001; Harm & Seidenberg, 1999, 2004; Perry, Ziegler & Zorzi, 2007; Plaut, McClelland, Seidenberg & Patterson, 1996; Zorzi, Houghton & Butterworth, 1998). Due to the importance of statistical structure in learning to read, some researchers suggest that reading acquisition is an exercise in statistical learning (e.g., Harm & Seidenberg, 2004; Sawi & Rueckl, 2019) and emphasize the importance of domain-general abilities in reading and learning to read.

Theories of learning regularities in a reading context

Because of the importance of the statistical structure of language in reading and reading acquisition, a recent line of research has explored the relationship between individuals' statistical learning (SL) ability and reading performance, as well as learning to read in both first and second languages. SL is a learning ability that extracts distributional properties from sensory input across time and space (Frost, Armstrong, Siegelman & Christiansen, 2015), and it is involved in various domains of cognitive processes (e.g., Baldwin, Andersson, Saffran, & Meyer, 2008; Cleeremans, 1993; Pacton, Perruchet, Fayol & Cleeremans, 2001; Saffran, Johnson, Aslin, & Newport, 1999; Saffran, Newport, & Aslin, 1996; Thiessen, 2010). A typical task used to study SL is the paradigm developed by Saffran and others (e.g., Saffran, Aslin & Newport, 1996), in which participants listen to a stream of syllables that varies in transitional probabilities and are tested on whether their responses are shaped by these probabilities. Other measures of SL include serial reaction time tasks (SRTT) and artificial grammar learning tasks (AGL) which also involve detecting probabilities in sequences.

The relationship between statistical learning and learning to read has been supported by evidence from two aspects. First, direct evidence on the relationship between SL and learning to read has been provided by studies in L2 learning (e.g., Frost, Siegelman, Narkiss & Afek, 2013; Wu et al., 2012). Second, empirical evidence has shown that SL is related to reading performance or literacy-related abilities (e.g., Arciuli & Simpson, 2012; Daltrozzo et al., 2017; Qi et al., 2019; Spencer et al., 2015; van der Kleij et al., 2018).

Although many studies have shown that better performance on SL tasks is associated with better performance on learning to read and greater language proficiency in general, other studies didn't find such relationships (e.g., Kelly, Griffiths & Frith, 2002; Schmalz et al., 2018; Spencer et al., 2015). The reason for these inconsistent findings is probably that both reading and SL are componential and involve different processes. For example, the DRC model (Coltheart, 2006; Coltheart et al., 2001) suggests that different approaches are used when reading regular and irregular words, and the triangle model (e.g., Harm & Seidenberg, 2004) showed that word reading can be achieved by different strategies through O-P or O-S-P pathways. Therefore, different reading tasks may measure different types of reading abilities. SL has also been found to have multiple facets and is not a general unified skill (e.g., Armstrong et al., 2017; Sawi & Rueckl, 2019; Schmalz et al., 2018; Siegelman et al., 2017; van der Kleij et al., 2018). A large body of evidence has shown the modality specificity of SL, with very limited transfer and very low correlation between visual and auditory SL (Conway & Christiansen, 2005, 2006; Emberson, Conway & Christiansen, 2011; Redington & Chater, 1996; Siegelman & Frost, 2015; Tunney & Altmann, 1999). SL ability is also found to be specific to the type of statistics and the task (e.g., Conway & Christiansen, 2006; Gómez & Maye, 2005; Henderson & Warmington, 2017; Johansson, 2009; Misyak & Christiansen, 2012; Schmalz et al., 2018; Siegelman & Frost, 2015). Therefore, whether the relationship between reading or reading acquisition and SL can be found is probably affected by the types of tasks that are used to measure the two capacities, as well as individual differences in doing these tasks (Sawi & Rueckl, 2019).

Another theory that relates reading acquisition to domain-general learning is the multiple memory systems model of declarative and procedural learning (DP model; e.g., Ullman, 2004; Ullman et al., 1997; Ullman & Pierpont, 2005). The DP model in language is derived from the theory of multiple memory systems proposed by Squire (1992, 2004). Squire proposed two general types of memory systems – declarative memory which stores conscious memories of facts and events, and procedural memory which stores unconscious memories of skills, habits, priming, etc. Declarative learning occurs with consciousness in a short period of time and is used to learn arbitrary relationships. It is domain general (Eichenbaum & Cohen, 2001; Squire & Knowlton, 2000). Procedural learning requires repetition and practice over time, without direct consciousness, and is used to learn complex relationships of sequences. It is domain specific (Squire & Knowlton, 2000; Ullman, 2004). In regards to language learning, declarative learning underlies the learning of the idiosyncratic form of words and other word-specific knowledge, whereas procedural learning system is used to learn "mental grammar" such as the regular past tense in English. Since O-P mapping contains more regularities compared to O-S mapping, O-P learning relies more on procedural learning compared to O-S learning (Ullman & Pierpont, 2005).

The procedural deficit hypothesis (PDH; Nicolson & Fawcett, 2007; Ullman, 2004) based on the DP model suggests that impaired procedural learning may be one cause of dyslexia.

Dyslexic individuals suffer from a specific impairment of phonological processing (Bradley & Bryant, 1983; Snowling, 1987, 2000; Vellutino, 1979). According to an early 'phonological' theory, this impairment results in difficulty with the learning of grapheme – phoneme rules (e.g., Snowling, 1987, 2000). Because procedural learning is used to learn regularities according to the DP model and there is more systematicity in O-P correspondences compared to O-S correspondences, a deficit in procedural learning may underlie the impairment in learning O-P correspondences. The PDH has been supported by studies which directly explore the relationship between procedural learning and dyslexia (e.g., Cassar & Treiman, 1997; Folia et al., 2008; Gombert, 2003; Menghini, Hagberg, Caltagirone, Petrosini & Vicari, 2006; Nigro, Jiménez-Fernández, Simpson & Defior, 2016). For example, Pavlidou, Louise Kelly and Williams (2010) used an artificial grammar to measure procedural learning, and dyslexic children performed at chance level on this task which was lower than their normal developing peers. Vicari et al. (2003) measured procedural learning with a serial reaction time task (Nissen & Bullemer, 1987) and reported that dyslexic children showed a reduced learning rate in this task as well.

Although the procedural deficit hypothesis was supported by many studies which have shown that on average, individuals with dyslexia have worse procedural learning abilities than controls (see Lum, Ullman & Conti-Ramsden, 2013), the same findings were not always replicated (Inácio et al., 2018; Kelly, Griffiths, & Frith, 2002; Rüsseler, Gerth, & Münte, 2006; Staels & Van den Broeck, 2015; Vakil, Lowe & Goldfus, 2015; Waber et al., 2003; West, Vadillo, Shanks & Hulme, 2018). For instance, using similar stimuli with Pavlidou et al. (2009, 2010) but in a different artificial grammar learning task, Inácio et al. (2018) failed to find impaired implicit learning in dyslexic children. They argued that extending the acquisition phase and including consolidation processes may be the reasons why their dyslexic group achieved much better performance. Some other studies did not find differences between dyslexic and control groups in overall performance (e.g., Howard, Howard, Japikse & Eden, 2006; Nigro et al., 2016; van der Kleij, Groen, Segers & Verhoeven, 2019). Henderson and Warmington (2017) also pointed out that whether implicit sequential learning predicts reading deficits depends on the task.

A recent study by West et al. (2018) reported low reliability between different procedural memory measures and failed to find a correlation between any procedural memory measure they used and literacy performance in a large sample of normal developing children. They questioned the procedural deficit hypothesis and suggested that more reliable measures of procedural memory were needed (but see Conway, Arciuli, Lum & Ullman, 2019 for the weakness of the study and the reply by West, Vadillo, Shanks & Hulme, 2019). Therefore, the reason why the evidence for the procedural deficit hypothesis is mixed may be similar to that for SL and reading acquisition. Performance on procedural learning tasks can reflect more than one element, with perceptual, cognitive, and motor processes all being involved, and when using different procedural learning tasks, they are not measuring exactly the same ability, which is probably why low reliability was found among different tasks.

Another possibility for the mixed evidence is that the process of learning regularities may be complex and involve~~s~~ more than one system. This is not well studied in the reading literature, but it is already explored in visual category learning which is purely about categorizing different patterns. In visual category learning literature, the COVIS (the competition between verbal and implicit systems) model suggests that individuals can use two systems to learn regularities. One is a *reflective* learning system which acquires rule-based regularities with conscious awareness, and the other is a *reflexive* learning system which acquires probabilistic regularities and integrates information from multiple dimensions without awareness of the specific regularities. Therefore, using the COVIS model to investigate the process of learning regularities in reading acquisition may improve our understanding about the underlying mechanisms.

Reflective/reflexive learning and learning to read

The contrast of reflective and reflexive learning was proposed by Ashby et al. (1998) and Ashby and Waldron (1999) in the field of category learning. In this COVIS model, the two systems compete with each other: reflective learning is an explicit, hypothesis-testing system dependent on working memory and executive attention to discover the rules for explicit classification (i.e., rule-based category learning), and the reflexive learning system requires information integration, is not consciously penetrable, and operates by associating perception with actions that lead to reinforcement via feedback.

The COVIS model in visual and speech category learning provides valuable insight into how regularities are acquired by the reflective and reflexive learning systems (e.g., Ashby & Maddox, 2011; Ashby & Valentin, 2005; Chandrasekaran, Yi & Maddox, 2014). It suggests that reflective and reflexive learning are two independent systems. The biochemical mechanism of reflexive learning determines that if correct responses are followed immediately by a reward signal, the corresponding synapses can be strengthened, so, reflexive learning requires presenting feedback in time. On the other hand, reflective learning relies on working memory and executive attention, and the learning occurs with conscious awareness, so it is not as sensitive to the timing of feedback but will be affected by a secondary task which requires the same cognitive resources. Since reflexive learning does not involve working memory and attention, it should not be affected by this type of secondary task. By manipulating tasks, behavioral empirical studies have confirmed that reflective and reflexive learning systems are independent from each other. For example, studies of Ashby, Isen and Turken (1999) and Ashby, Queller and Berretty (1999) showed that reflexive learning was disrupted if no feedback was provided in a visual category task. Maddox, Ashby and Bohil (2003) compared the performance of reflexive learning with immediate and delayed feedback in a categorization task and found that delayed feedback significantly disrupted reflexive learning performance. None of these studies found any impact on reflective learning by manipulating feedback. These studies confirmed that providing feedback and the timing of feedback are important to reflexive learning but not to reflective learning. In another study, Waldron and Ashby (2001) used a single-task condition which was a categorization task and a dual-task condition which required participants to memorize the physical size and numerical value of numbers while doing the categorization task. They found that for the reflective learning participants needed more training in the dual-task condition compared to participants in the single-task condition to achieve the same performance, but there was no such interference in the reflexive learning. This shows that a secondary task disrupted reflective learning severely, but not reflexive learning.

The COVIS model also stipulates that the two learning systems compete with each other, which has been supported by neural imaging studies showing that as activation in basal ganglia increases, activation in MTL decreases (e.g., Frank, O'Reilly & Curran, 2006; Moody, Bookheimer, Vanek, & Knowlton, 2004; Poldrack et al., 2001; Poldrack, Prabhakaran, Seger, & Gabrieli, 1999; Seger & Cincotta, 2005; Seger & Miller, 2010). Using a weather prediction task, Poldrack et al. (2001) found that the activation in MTL and caudate nucleus were negatively correlated. A similar pattern was also found by Seger and Cincotta (2005) in an imaging study ~~that~~; activation in the caudate correlated negatively with activation in hippocampus. Studies on patients with brain lesions also support the competition between the two learning systems. Frank et al. (2006) found that patients with damage in MTL showed better performance on a task involving the basal ganglia, and Moody et al. (2004) reported that during probabilistic classification category learning which ~~was~~ is a reflexive learning task, individuals with damage ~~in~~ to the basal ganglia due to Parkinson's disease used the MTL to a larger extent as compared to ~~the~~ a control group.

Therefore, in the current study I would like to address two major questions. Research in visual category learning suggests a possibility that learning regularities is heterogenous, and my first questions is whether learning to read, which also requires learning regularities between orthography, phonology, and semantics, involves reflective and/or reflexive learning. The second question is whether O-P and O-S learning relies on different mechanisms. As mentioned earlier in both computational modelling and the DP model, O-P mapping contains more regularities than O-S mapping, and this difference may lead to different usage of the learning systems. In this study I will use an artificial orthography learning (AOL) paradigm to measure participants' learning to read.

Learning to read and paradigm of artificial orthographies

Recently more and more studies have examined how adults learn to read using a new orthography of made-up words rather than observing them learning a natural language (e.g., Bitan & Karni, 2003, 2004; Deng et al., 2008; Deng et al., 2011; Havas et al., 2015; Merkx, Rastle & Davis, 2011; Moore et al., 2014; Rueckl & Dror, 1994; Taylor et al., 2011; Xue et al., 2006; Yoncheva et al., 2015). This methodology minimizes the impact of each individual's reading experience in the experiments, and the input statistics can be better controlled in artificial words than in real words selected from lexical databases of natural languages. Many studies have shown that AOL is a good paradigm to investigate the learning of O-P and O-S regularities with different statistical structure built into the orthography (e.g., O-P learning: Byrne, 1984; Deng et al., 2011; Taylor et al., 2011; Trudeau, 2006; O-S learning: Merks et al., 2011; Rueckl & Dror, 1994).

In a recent study, Zhao, Li, Elliott and Rueckl (2017) manipulated consistency in both O-P and O-S mappings in Chinese pseudowords and trained English-speaking participants with AOL tasks. In the training phase, participants were required to learn the artificial orthography by learning the meaning and pronunciation of each word, i.e., the O-S and O-P mappings. In this lexicon, each word was made by two Chinese characters that constituted two radicals. Consistent and inconsistent O-P and O-S mappings were created. In the consistent mappings, one of the radicals (semantic radical) in each word was related to the semantic category of the whole word's meaning, and the other radical (phonological radical) indicated the rhyme of the word. In this way, sublexical regularities of the O-P mapping and the O-S mapping were built into the orthography. After training, participants performed a categorization task to see whether they were able to generalize the regularity to novel words with the same structure. For O-P categorization, participants saw artificial words and decided how to pronounce their rhymes. In addition to the trained words, new words were created for this task, each of them contained a phonological radical which was same as the trained words and a novel radical that had never appeared before. If participants had learned the sublexical O-P regularities, they should not only know the rhyme for the trained words, but also be able to generalize the knowledge to new words that contained the same phonological radicals. Similarly, for O-S categorization, participants saw artificial words and decided on their semantic categories for both trained and new words. Findings of this study confirmed the results in previous research that individuals learned faster when statistical structure exists in O-P or O-S mappings, i.e., people learned consistent O-P and consistent O-S mappings faster and better than inconsistent mappings, and the consistent regularities can be generalized to novel words with similar structure.

This AOL paradigm has been used in our lab with different adjustments to examine the relationship between learning to read and statistical learning ability. We have found that performance on a visual statistical learning (VSL) task predicted learning both O-P and O-S mappings when the two mappings were consistent (Rueckl, Li, Brown & Zhao, 2016). VSL has been found to involve both the striatum and MTL systems (Turk-Browne et al., 2009). Therefore, given the robust relationship we have found between the AOL and VSL performance, this AOL task should be a good task to examine whether reflective and reflexive learning are engaged in learning to read.

The Current Study

The goal of this study is to investigate whether learning to read involves reflective and/or reflexive learning, and whether O-P and O-S learning relies on different mechanisms. In Experiment 1, I would like to use different manipulations to disrupt reflective or reflexive learning (Ashby et al., 1999; Maddox et al., 2003; Waldron & Ashby, 2001) on the AOL task. In Experiment 2, I would like to examine the relationship between individual differences in reflective and reflexive learning systems and learning in the AOL task.

The AOL task used in this study was similar to the one in Zhao et al. (2017) where both O-P and O-S mappings were consistent. In this task, sublexical regularities of O-P and O-S mappings were built into the orthography. Like in Zhao et al. (2017), each word was composed of two radicals, one of which provides information about the rhyme of the word's pronunciation and the other providing information about the word's semantic category. In the training task, participants learned each word's meaning and pronunciation by selecting the correct picture indicating the meaning of the word on the screen or selecting its correct pronunciation from two options. They would receive feedback telling them whether their selection was correct or not after each trial, and the correct answer was also shown or heard with the feedback. After training, a categorization task which was same as the categorization task in Zhao et al. (2017) was conducted to see whether participants were able to generalize the sublexical regularity to novel words with the same structure. More details about this task are presented in the method section.

In order to find out whether the reflective and reflexive learning systems were used in learning to read, different manipulations were used in the AOL task in Experiment 1 to disrupt reflective learning and reflexive learning~~,~~ respectively, and to see whether participants' performance was impaired in these two conditions compared to the baseline condition. The manipulations – delaying feedback and conducting a concurrent task – were borrowed from studies on visual category learning (e.g., Maddox & Ashby, 2004). In Experiment 2, reflective and reflexive learning ability was directly measured through category learning tasks, and the relationship between the reflective/reflexive learning abilities and performance in AOL tasks was examined to further investigate the interactivity between the two learning systems in AOL tasks.

Experiment 1

The aim of Experiment 1 was to examine whether reflective and reflexive learning are engaged in learning to read. Specifically, I wanted to see if reflective and reflexive learning would be disrupted by two different manipulations in an artificial lexicon learning (AOL) task. One manipulation would be to delay feedback to affect dopamine release and disrupt reflexive learning and another manipulation would be to introduce a concurrent task to reduce processing resources and disrupt reflective leaning. If performance on the AOL tasks with reflexive or reflective learning disruption is worse than the performance in a baseline condition, it would support the idea that one or the other type of learning is engaged in the AOL tasks.

To achieve the aim, three between-subject conditions were used in the AOL task: 1) an immediate feedback condition (baseline); 2) a delayed feedback condition (DF condition); and 3) an immediate feedback with a concurrent task condition (CT condition).

The DF condition was used to disrupt the effectiveness of reflexive learning. As explained in the introduction, the effect of feedback for reflexive learning is time sensitive and learning will not benefit from delayed feedback (Gamble & Koch, 1987; MacDermott et al., 1986). Evidence has shown that delayed feedback leads to lower accuracy in an information-integration (reflexive) category-learning task but has no effect on the response accuracy in a rule-based (reflective) category-learning task (Maddox, Ashby & Bohil, 2003). Some studies of visual category learning have reported that feedback delays of 2.5 seconds or longer impair reflexive learning, whereas delays of even 10 seconds have no effect on reflective learning (Dunn, Newell & Kalish, 2012; Maddox et al., 2003; Maddox & Ing, 2005). Considering that the AOL task is rather long even with immediate feedback, I used a delay near the short end of this range (3 seconds) in the DF condition.

Similarly, the CT condition was used to disrupt the effectiveness of reflective learning. Reflective learning depends on executive function, so a concurrent task adopted from Waldron and Ashby (2001) which required working memory and executive attention would use the same resources and have a negative impact on reflective learning (Maddox & Ashby, 2004). Because reflexive learning does not depend on executive function, it would not be affected by this concurrent task (Waldron & Ashby, 2001). If accuracy of responses in the CT condition is lower compared to the baseline condition, it would suggest that reflective learning was used in learning to read in the AOL tasks.

Method

Participants.

Seventy-one undergraduate students from University of Connecticut participated in the Experiment 1. Among them, 24 were in baseline condition, 22 in DF condition, and 25 in CT condition. All participants were native speakers of American English, and none had knowledge of Chinese or Japanese.

The number of participants in each group was determined based on previous research using the AOL (Zhao et al., 2017) and visual category learning (Maddox, Ashby, Ing & Pickering, 2004; Maddox, Filoteo, Hejl, & Ing, 2004; Zeithamova & Maddox, 2007) tasks. In each of these studies, 20 to 25 participants per condition provided enough power to yield significant learning effects.

Materials.

The materials used in the training task in present study were same to those in Zhao et al. (2017). Ten Chinese simple radicals (“士”, “中”, “下”, “刀”, “山”, “小”, “大”, “儿”, “广”, and “攵”) were used to create artificial lexicons. They were divided into two sets of equal number. One set was used as semantic radicals and the other as phonological radicals; the set assigned as semantic/phonological radical was counterbalanced. The two groups of radicals were combined to make artificial words. Each “word” contained one semantic radical and one phonological radical, creating 25 (5×5) words. All words were formed with one radical on top and one on the bottom. For the O-S mapping, each word was associated with a specific meaning. Five semantic categories (animal, body part, fruit, furniture, and clothing) were used, each of which corresponded to one semantic radical, therefore, producing five words in each semantic category. In each of the semantic categories, five high-rank exemplars were chosen based on the category norms by Battig and Montague (1969) and were presented to participants with black and white pictures from Snodgrass and Vanderwart's (1980) picture database. For the O-P mapping, each word was associated with a specific pronunciation. Each of the five rhymes (/-eɪs/, /-ɜrb/, /-aɪv/, /-æd/, and /-ʌk/) corresponded to one phonological radical. Therefore, each rhyme corresponded to five words. The positions of semantic and phonological radicals were counterbalanced between subjects; for example, when 士 was a semantic radical and 小 was a phonological radical, some participants learned the word 赤 and some others learned 尘 .

A categorization task was conducted after training to test whether the participants could generalize the regularities to novel words with a similar structure. For O-S mapping, a set of 25 transfer words were created by combining the five semantic radicals with five novel radicals ("子", "爪", "手", "尸", and "戈"). Therefore, the semantic transfer words shared the same semantic radicals with the trained words, but the other part of the word was a novel radical rather than an original phonological radical. For the O-P mapping, a set of 25 transfer words were created in the similar way, this time with five novel radicals combined with the five phonological radicals. Therefore, the phonological transfer words shared the same phonological radicals with the trained words, but the other part of the word was a novel radical rather than an original semantic radical.

Procedure.

Participants first completed a training task and then a categorization task. Each participant learned O-P and O-S mappings, respectively. In an O-S training trial, participants saw an artificial word on the screen with two pictures below it, and they were asked to judge which of the two pictures matched the meaning of the word. After making responses by clicking on the picture, they received feedback telling them whether they were correct or not, and then they saw the word with the picture of the correct meaning. In an O-P training trial, participants saw each artificial word and two boxes with numbers 1 and 2 on them. They did not see pictures. Participants would hear two pronunciations and be asked to select which one was the correct pronunciation of the word by clicking on box 1 or 2. When given feedback, they would see the word again and the correct box and would hear the correct pronunciation.

As mentioned above, there were three conditions: baseline, DF and CT. In the baseline condition, participants received immediate feedback and had a blank interval for 3.5s before the start of the next trial. In the DF condition, the feedback was delayed for 3 seconds and the next trial started after 500ms. In the CT condition, at the beginning of each trial, participants saw two numbers of different sizes and values presented on the screen, one on the left and one on the right, for 200ms. They then saw the word with the two pictures (in O-S training) or with the two boxes and heard two pronunciations (in O-P training) and, after making their response, they received immediate feedback just like in the baseline condition. They would then be asked to report either the larger number in value or in size according to the question shown on the screen by clicking one of the two boxes which appeared in the same position as the two numbers at the beginning of the trial. The next trial would start after 3.5 s.

The training task included six O-S training blocks and six O-P training blocks with 25 trials in each block. The two types of training alternated across blocks, e.g., if the participant's first block was O-P training, then the second block was O-S training, and the third block was again O-P training, etc. The order (which type of training occurred first) was counterbalanced across participants.

In the categorization task, for O-S mapping, in each trial an artificial word was presented with an English word below it (e.g., ANIMAL), and participants had to judge whether the meaning of the artificial word belonged to the semantic category as English word. To test whether participants learned the structure of the mapping, there were also 25 transfer words, each of which had a semantic radical as in the trained words and a novel radical which had never appeared before. For O-P mapping, an artificial word was presented with an English word below it (e.g., RACE), and participants had to judge whether the artificial word rhymed with the English word. To test whether participants truly acquired the structure of the mapping, there were also 25 transfer words, each of which had a phonological radical as in the trained words and a novel radical which had never appeared before. No feedback was provided in the categorization task. There was one O-S and one O-P categorization block in this task and the order of the two blocks was counterbalanced between participants.

The experiment took about sixty minutes.

Results

Generalized linear mixed models using the lme4 package (version 1.1-13) in R (Version 3.4.1; Bates et al., 2017) were used to compare learning in the three conditions. For both training and categorization tasks, the performance in DF and CT conditions was compared to that in the baseline condition. In addition, performance in DF and CT conditions was compared to each other. The reason that I compared the three conditions in pairs rather than in one model is that a single model cannot separate the two levels of DF and CT. That is to say, if using one model, when condition was added to the model, both DF and CT would be added at same time; therefore, the contribution of the two manipulations to the improvement of the model fit could not be separated using only one model. The research question of this experiment was whether reflexive and reflective learning was engaged in learning to read, and comparisons between baseline and DF and between baseline and CT were sufficient to answer this question. However, it was also interesting to examine which manipulation had stronger impact on learning by comparing CT and DF conditions directly, so this DF-CT comparison was also included, although it was not a part of my hypothesis.

The models included random effects for participants, but not random effects for items, because counterbalancing was already conducted on the word structure (position of radicals and which radicals were used as phonological/semantic radicals) to minimize the impact of each particular item on learning. Barr, Levy, Scheepers and Tily (2013) showed that maximal random effects structures minimize false alarm rates without substantial loss of power using Monte Carlo simulation. Therefore, in the analysis I tried to keep the random effects maximal, such that I included all of the factors that could hypothetically vary across individuals.

Results were presented for the training and categorization tasks.

Training task

Figure 1 shows the mean accuracy for each block in baseline, DF, and CT conditions for O-P and O-S mappings separately. As can be seen in the figure, accuracy of performance gradually improved with more blocks, and the overall learning in O-S mapping was better than in O-P mapping. Logistic growth curve models were structured according to Mirman (2014) to analyze the training data. Performance of learning was first modeled with fixed effects of block (linear and quadratic orthogonal polynomials) (Model 1), then fixed effects of mapping and interactions between mapping and block were added (Model 2), and then fixed effects of condition and interactions between condition and other variables were added (Model 3a, 3b, 3c), as shown in Table 1. For random effects, Model 1 only included linear and quadratic terms of blocks, and Model 2, 3a, 3b and 3c included both linear and quadratic terms of blocks and mapping.

Model 1 included the fixed and random effects of block (linear and quadratic orthogonal polynomials), and it showed that performance improved over blocks, which reflected the pattern in Figure 1. Model 2 added the fixed effects of mapping and interaction between mapping and block, as well as random effects of mapping. Mapping was encoded using effects coding, i.e., O-P = -0.5, O-S = 0.5. In the fixed effects of linear term of block, mapping and interaction between block and mapping were all significant, and adding the fixed effects of mapping and the interaction between mapping and block and the random effect of mapping significantly improved the model fit, $\chi^2(9) = 418.55$, $p < .001$. The results confirmed what Figure 1 suggests; learning improved with more blocks, and learning in O-S mapping was better than in O-P mapping. In addition, the interaction between mapping and block showed that with more blocks, learning the O-S mapping improved more than in the O-P mapping. The results were also consistent with previous experiments in our lab using the AOL tasks (Rueckl et al., 2016).

Model 3a, 3b and 3c examined the differences between DF and baseline, CT and baseline, and DF and CT conditions, respectively. Effects coding was used in each comparison. For the baseline-DF comparison, a new variable was created with baseline= -0.5, DF= 0.5, and CT = 0; for the baseline-CT comparison, a new variable with baseline= -0.5, DF = 0, and CT= 0.5; and for the DF-CT comparison, another variable was added with baseline= 0, DF = -0.5, and CT= 0.5.

Model 3a added the fixed effects of the variable baseline-DF as well as its interaction with mapping and block. The inclusion of these factors significantly improved the model fit, with $\chi^2(4) = 9.61$, $p < .05$. In the model, baseline-DF was not significant, but the interaction between condition and mapping was. Although the main effect of condition was not significant, the interaction indicated a nonsignificant trend that the overall performance in DF was better than that in the baseline in the O-P mapping (mean difference between baseline and DF was -.02), and a nonsignificant trend of the overall performance in DF was worse than that in baseline in O-S mapping (mean difference between baseline and DF was .01). These trends are also shown in Figure 1.

Model 3b added the fixed effects of the variable baseline-CT as well as its interaction with mapping and block to Model 3. Adding the baseline-CT contrast significantly improved the model fit with $\chi^2(4) = 24.55$, $p < .001$. In the model, baseline-CT was significant, as well as the interaction between baseline-CT and mapping. The significant fixed effects indicated that, overall, performance in CT condition was worse than in the baseline, and the difference between performance in the two conditions was larger in the O-S mapping than in the O-P mapping.

Model 3c added the fixed effects of the variable DF-CT as well as its interaction with mapping and block. Adding the DF-CT contrast significantly improved the model fit, with $\chi^2(4) = 13.51$, $p < .01$. The model showed that DF-CT and its interaction with block were significant. The significant fixed effects indicated that overall, performance in CT condition was worse than in DF condition, and the difference between performance in the two conditions became larger with more blocks. This was consistent with Model 3a and 3b which showed that overall performance in the baseline and DF condition was about the same, but performance in CT was worse than that in the baseline.

In sum, overall performance improved with more training blocks, and O-S learning was better than O-P learning in general, which was consistent with previous studies using similar AOL tasks (Rueckl et al., 2016; Zhao et al., 2017). Critically, performance in the CT condition was significantly worse than that in the baseline condition in the training task, and O-S learning was more affected by the concurrent task compared to O-P learning. In the DF condition, the delayed feedback affected O-P and O-S learning differentially, with performance in DF a little better than in the baseline for O-P learning, and performance in DF a little worse than in the baseline for O-S learning.

Categorization task

Figure 2 shows the mean accuracy for categorization task in baseline, DF, and CT conditions for O-P and O-S mappings separately. As can be seen in the figure, there was no difference between performance on trained and transfer words. Mixed effects logistic regression models were structured according to Baayen, Davidson and Bates (2008) to analyze the categorization data. Accuracy was first modeled with fixed effects of word type (Model 1), then fixed effects of mapping and interaction between mapping and word type were added (Model 2), and then fixed effects of condition and interactions between condition and other variables were added (Model 3a, 3b, 3c), as shown in Table 2. For random effects, Model 1 only included word type, and Model 2, 3a, 3b and 3c included both word type and mapping.

Model 1 included the fixed and random effects of word type. As in Figure 2, word type was not significant in predicting categorization performance. Model 2 added the fixed effects of mapping and interaction between mapping and word type, as well as random effects of mapping. Adding mapping terms significantly improved the model fit, $\chi^2(5) = 366.90$, $p < .001$. The fixed effect of mapping was significant, showing that performance in O-S mapping was better than that in O-P mapping, as shown in Figure 2.

Applying the same strategy used in the analysis of the training data, Models 3a, 3b and 3c were built to examine the difference between DF and baseline, CT and baseline, and DF and

CT conditions, respectively. Model 3a showed that adding the baseline-DF contrast marginally improved the model fit, with $\chi^2(3) = 7.27$, $p = .06$. In the model, baseline-DF was not significant, but the interaction between condition and mapping was marginally significant. This pattern was very similar to that in the training task in which the fixed effect of condition was insignificant but the interaction between condition and mapping was. The interaction showed a nonsignificant trend that performance in DF condition was better than in the baseline condition, and the difference in O-P categorization was larger than that in O-S categorization.

Model 3b showed that adding the baseline-CT contrast significantly improved the model fit, with $\chi^2(3) = 13.85$, $p < .01$. In the model, baseline-CT was significant, as well as the interaction between baseline-CT and mapping. The pattern was the same as in training task, showing that overall performance in baseline was better than that in CT condition, and the difference between performance in the two conditions was larger in O-S mapping than in O-P mapping.

Model 3c showed that adding DF-CT contrast significantly improved the model fit, with $\chi^2(3) = 20.23$, $p < .001$. In the model, DF-CT was significant. This was consistent with the pattern in training task, both showing that performance in CT was worse than that in DF condition.

Taken together, the patterns found in training and categorization tasks were very similar to each other. The concurrent task significantly impaired AOL performance but delayed feedback did not.

Discussion for Experiment 1

Experiment 1 clearly showed that learning improved over blocks and learning the O-S mappings was always better than learning the O-P mappings. This was as expected, for the same pattern has also been found in other learning experiments using the same AOL tasks in our lab (e.g., Rueckl et al., 2016; Zhao et al., 2017).

The comparison between the Baseline and CT conditions confirmed that the concurrent task had a stronger impact than delayed feedback on learning. Specifically, the concurrent task had a significant negative impact on learning, accuracy was higher in the Baseline condition than in the CT in both the training and categorization task. On the other hand, delayed feedback only had a weak impact on AOL learning, for in both training and categorization, only the interaction between Base-DF and mapping was significant, but the main fixed effect was not. When the interaction was further examined, performance in DF condition was slightly better than that in Baseline, which was unexpected. A possible reason for this unexpected trend may be attention. In the delayed feedback condition, the 3 seconds delay may help participants get more focused while waiting. Reflexive learning does not require attention since it is essentially automatic with appropriate feedback, but reflective learning uses logical reasoning and depends on executive attention (Maddox & Ashby, 2004). Therefore, although the delayed feedback weakened reflexive learning, the enhanced attention might get reflective learning more engaged. If this was true, the trend that performance in DF condition was slightly better than that in Baseline was understandable.

One interpretation of the results in Experiment 1 was that AOL learning depends on reflective but not on reflexive learning because the concurrent task (hypothesized to disrupt reflective learning) lowered performance relative to the Baseline condition, but delayed feedback (hypothesized to disrupt reflexive learning) did not. In addition, performance with the concurrent task was worse than with delayed feedback, which also suggests that the reflective learning was engaged in the AOL task. An alternative interpretation involves competition between the two learning systems (e.g., Poldrack et al., 2001)—specifically, the possibility that when one learning system is disrupted, the other becomes more engaged. According to this interpretation, both learning systems supported learning in the Baseline condition and (based on the direct comparison between the DF and CT conditions) reflective learning played a larger role. When reflexive learning was disrupted by delayed feedback, participants relied more on reflective learning in the tasks and overall performance was not affected much due to this compensation. In contrast, when reflective learning was disrupted by the demands of the concurrent task, reflexive learning may have been more engaged, it was not sufficiently effective to compensate the loss from disrupted reflective learning. The results of Experiment 1 do not clearly support one interpretation over the other. However, both interpretations imply that reflective learning contributed *more* to AOL performance in Experiment 1.

The significant interaction between mapping and Base-CT showed that O-S learning was affected more by the concurrent task than O-P learning. There are also two possible interpretations of this finding. One is that learning regularities in the O-S mapping is more dependent on reflective learning. The other is that the interaction is due to the poor overall performance in O-P mapping. Since learning in O-S mapping was always better than learning in O-P mapping, it may be easier to see the impact on performance in O-S mapping when disrupting reflective learning. Again, Experiment 1 cannot show which interpretation was better.

Taken together, results from Experiment 1 indicated that reflective learning was engaged in the AOL tasks. Whether reflexive learning was engaged and whether there was a competition between the two learning systems were still unclear. Also, whether O-P and O-S learning relied on the two learning systems similarly or not was not clearly revealed by Experiment 1. Therefore, Experiment 2 directly measured reflective and reflexive learning, and whether participants' reflective and reflexive learning ability can predict performance in AOL tasks was the main question to investigate.

Experiment 2

The main purpose of Experiment 2 was to examine the relationship between reflective/reflexive learning ability and individual differences in learning to read. It was based on the results from Experiment 1, which suggested that reflective learning was engaged in learning in the AOL task, as shown by the finding that-the performance in both the training and categorization tasks in AOL was worse when reflective learning was disrupted. However, the engagement of reflexive learning was not as clear. Delaying feedback in training did not affect the overall performance in either training or categorization, but the interaction between condition and mapping in both tasks suggested that reflexive learning may be engaged differentially in the performance of O-P and O-S mappings. Therefore, Experiment 2 directly measured reflexive and reflective learning to examine whether the two types of learning can predict AOL performance.

The measure for reflective and reflexive learning was borrowed from a task used by DeCaro et al. (2008). In this task, there are two types of regularities to learn. For one regularity, visual stimuli could be categorized by a unidimensional rule (rule-based categorization), and for the other, visual stimuli could only be categorized by integrating information from multidimensions (information-integration learning), making this categorization regularity very hard to state verbally. Based on the correspondences between the neural circuits used and the tasks, as mentioned earlier (Nomuna & Reber, 2008), this task (measuring a unidimensional rule-based learning and a multidimensional information-integration learning) was used to measure reflective learning and reflexive learning, respectively.

In addition to the reflective/reflexive tasks, we wanted to examine visual statistical learning (VSL) which has been found to be a strong predictor of the AOL performance (Rueckl et al., 2016). This suggests a positive relationship between statistical learning (SL) and learning to read. However, the complex mechanism of SL as a componential ability requires more evidence to understand this relationship. Therefore, a secondary purpose of the Experiment 2 was to discover whether this experiment replicated the pattern and then examine the relationship between VSL and reflective/reflexive learning, to further investigate the mechanism of SL. Frost et al. (2013) found that VSL performance was not correlated with several general cognitive measures, suggesting that VSL was an independent cognitive capability. In the current study to ensure that performance in the VSL task was not confounded with these general cognitive abilities, an IQ task and a working memory task were included.

Another purpose of Experiment 2 to investigate the relationship between individual differences in reflective and reflexive learning. DeCaro et al. (2008) found that individuals who performed better on working memory tasks also performed better in the reflective learning task but worse in the reflexive learning task. Working memory and reflective learning share common brain regions, so it is reasonable that the two had a positive relationship, but it was surprising that reflexive learning was negatively related to working memory. DeCaro et al. argued that people who performed better on reflective learning may rely on complex computational processes when learning, but these processes are not necessarily optimal for the reflexive learning task. However, Kalish, Newell and Dunn (2017) found that participants with higher working memory capacity tended to perform better in both of the category learning tasks, regardless of the structure of categories. In light of these conflicting results, I chose to include a working memory task in the present experiment to examine this relationship further.

Method

Participants.

Seventy-three undergraduate students from University of Connecticut participated in the study. All of them were native speakers of American English, and none of them had knowledge of Chinese or Japanese.

Measures

Artificial lexicon learning task. This was the same as the task used in baseline condition in Experiment 1 except that there were five blocks for each mapping rather than six in the current experiment due to the limited time to finish all the tasks.

Visual category learning task (Reflective and reflexive learning measure). A visual categorization task from DeCaro et al. (2008) was used to measure reflective and reflexive learning. Participants viewed a stimulus on the screen and chose whether this stimulus belonged to category A or B, and then received feedback about whether their answer was correct or not. There were two blocks for reflective learning and two blocks for reflexive learning, with 200 trials in each block.

As in DeCaro et al. (2008), the stimuli were adapted from Waldron and Ashby (2001): sixteen stimuli were used that were a combination of four dimensions: background color (yellow or blue), symbol shape (circle or square), symbol color (red or green), and number of symbols (1 or 2). Rule-based regularities were one-dimensional (e.g., a circle corresponded to category A, and a square corresponded to category B). For the two rule-based regularity blocks, the selected dimensions were symbol color and symbol shape. Information-integrating regularities involved three dimensions for each of the two blocks. A value of -1 or +1 was randomly assigned to the two levels of each dimension (e.g., for the dimension of color, green = -1 and red = +1). The rule for categorization was that if the sum of the values of the three dimensions was larger than zero, this stimulus belonged to category A, otherwise belonged to category B (Waldron & Ashby, 2001). For example, in a reflective learning task, if the symbol color was the rule-based regularity, then all stimuli with green symbols belonged to category A and all stimuli with red symbols belonged to category B (see Figure 3 upper panel). In a reflexive learning task, if shape, symbol color, and the number of symbols were the three dimensions determining the category of a stimulus, then if value (shape) + value (color) + value (number) > 0, the item belonged to category A, otherwise category B (see Figure 3 lower panel). The dependent variable of the task was the number of trials taken to learn categorization rules to get eight correct trials in a row.

Four orders of the blocks (two reflective learning and two reflexive learning) were counterbalanced between participants: 1212, 1221, 2112, 2121 in which 1 indicated reflective learning and 2 indicated reflexive learning task. At the beginning of the task, participants were asked to categorize the stimulus on the screen as either category A or B in each trial. They were told not to deliberate too long when making their decision, and feedback would be provided after they made their choice. No information about regularities in the stimuli was mentioned in the instructions.

Working memory task. One of the most commonly used individual difference measure of working memory is the sentence span task by Daneman and Carpenter (Daneman & Carpenter, 1980; Turner & Engle, 1989). The measure used in Experiment 2 was a listening version of this task adopted from Kuperman and Van Dyke (2011). This listening version avoided individual differences in reading ability and was therefore a more suitable choice than Daneman and Carpenter’s task for this experiment. Participants listened to the recording of several sets of sentences which increased in number of words per sentence. After listening to the recording of each sentence, they were required to judge whether the sentence was true or false, and to memorize the last word of the sentence. After finishing the set of sentences, they were asked to recall the last words of each of the sentences. There was a practice set with three sentences, and in the formal task the length of the set started with four sentences, then gradually grew to six sentences, so the task became increasingly more difficult with more sentences per set.

VSL task. The VSL task was adopted from Siegelman, Bogaerts and Frost (2016). In the familiarization phase, sixteen distinct complex shapes were presented to participants on the screen one at a time in a consecutive stream for about 10 minutes. The stream of shapes included eight triplets, each of which contained three shapes in a fixed order. Each triplet was repeated 24 times in random order and the same triplet was never repeated twice in a row. Participants were required to watch the stream in the familiarization phase, but they were not given any information about the triplets or the order of the stimuli. After the 10-minutes familiarization task, participants were given two tests. The first test, which had 34 questions, was a pattern recognition test. Participants selected one triplet or pair that seemed more familiar from two-alternative or four-alternative forced-choice options. The second test had eight questions and was a pattern completion test in which participants had to complete a familiar pair or triplet by selecting one appropriate shape from three options.

Matrix Reasoning task. The matrix reasoning task was a subtest from WASI Performance IQ subtests which measured participants' nonverbal IQ.

Procedure.

The AOL task, VSL task, matrix reasoning task, visual category learning task, and working memory task were conducted in this order. The whole experiment took about 2 hours.

Results

The relationship between performance of reflective/reflexive learning and AOL tasks was examined. As in Experiment 1, training and categorization in the AOL tasks were analyzed separately. Then the two secondary analyses on VSL and working memory were performed.

In the visual category learning task there were two blocks for reflective learning and two blocks for reflexive learning. For each block, the number of trials to reach eight consecutive correct trials was recorded, and performance for each kind of learning was measured by the mean of the two blocks. Figure 4 shows that the distribution of reflective learning scores was highly skewed, with many participants learned the regularities in only a few trials. Therefore, the scores for both learning were log transformed and the distributions of these scores are shown in Figure 5. Figure 5 shows that in general, participants were better at learning the regularities on the reflective learning task than the reflective learning task. The scores were then standardized and multiplied by -1 when used in the models, in order to make reflective and reflexive learning scores consistent with other tasks, i.e., the higher the score, the better the performance.

The relationship between reflective/reflexive learning and the AOL tasks

Training task

Figure 6 shows the mean accuracy of learning for each block for O-P and O-S mappings in the AOL training task. Accuracy performance improved with more training blocks, and performance in O-S mapping was better than in O-P mapping.

Models 1 and 2 were structured in the same way as in Experiment 1, i.e., learning accuracy in the AOL task was first modeled using growth curve models, with random and fixed effects of block (linear and quadratic orthogonal polynomials) in Model 1. Model 2 added the fixed effects of mapping and the interaction between mapping and block, as well as random effects of mapping. Model 3a and 3b were built based on Model 2 to examine the impact of reflective/reflexive learning on performance on the AOL task, respectively. Model 4 included the fixed effects of both reflective and reflexive learning, as well as their interactions.

Table 3 shows the mixed effects models for the training task. Model 1 included the fixed and random effects of block (linear and quadratic orthogonal polynomials), and it showed that performance was better with more blocks. Model 2 added the fixed effects of mapping and the interaction between mapping and block, as well as random effects of mapping. The fixed effect of linear term of block, mapping, and the interaction between block and mapping were all significant, and adding the fixed effects of mapping, the interaction, and the random effect of mapping significantly improved the model fit, $\chi^2(9) = 288.93$, $p < .001$. The results of Model 1 and 2 confirmed the pattern shown in Figure 6 that learning improved with more blocks, learning in O-S mapping was in general better than in O-P mapping, and there was more improvement in learning the O-S mappings than learning the O-P mappings as training progressed. These results were consistent with Experiment 1 and the previous experiments in our lab using the AOL tasks (Zhao et al., 2017).

Figure 7 shows that the correlation between performance in training and reflective/reflexive learning were moderate. According to the scatterplots, the reflexive learning seemed to have a stronger relationship with AOL training performance ($r = .35$) compared to reflective learning ($r = .29$). Model 3a and 3b were built to examine the impact of reflective and reflexive learning on AOL training tasks, respectively. Model 3a added fixed effects of reflective learning, interactions between reflective learning and mapping, and between reflective learning and block. Adding reflective learning fixed effects did not improve the model fit, $\chi^2(4) = 6.54$, $p = .16$. However, reflective learning was significant in Model 3a, and the interaction between reflective learning and linear term of block was marginally significant. Therefore, Model 3a showed mixed results, with a trend that participants with better reflective learning scores performed better in the training task, and with more blocks, the impact of reflective learning grew stronger.

Model 3b added reflexive learning and its interaction terms based on Model 2. Similar to Model 3a, adding reflexive learning related terms did not improve the model fit, $\chi^2(4) = 6.88$, $p = .14$, but in the model, the fixed effect of reflexive learning was significant. The results showed a trend which was similar to that in Model 3a, that participants with better reflexive learning scores performed better in the training task.

Model 4 included all the fixed effects in Model 3a and 3b. In addition, the interaction between reflective and reflexive learning, the interaction between reflective learning, reflexive learning and linear term of block, and the interaction between reflective learning, reflexive learning and mapping were also added to the model. The three-way interactions were added to the model for two reasons. First, a marginal significant interaction between reflective learning and linear term of block was found in Model 3a, and it is important to see whether reflexive learning ability can affect this interaction. Second, although no interaction between mapping and reflective/reflexive learning was found in Models 3a and 3b, in Experiment 1 there was a trend that disrupting reflective/reflexive learning affected the two mappings differentially, and it is crucial to measure this trend directly in Experiment 2. Model 4 did not improve Model 3a, Model 3b, or Model 2, $\chi^2(7) = 8.92$, $p = .26$, $\chi^2(7) = 8.58$, $p = .28$, and $\chi^2(11) = 15.46$, $p = .16$, respectively. However, similar to Models 3a and 3b, both reflective learning and reflexive learning were significant in Model 4. The interaction between reflective learning and block was marginally significant, as well as the interaction between reflective and reflexive learning. The results of Model 4 showed similar trends as Models 3a and 3b, in that participants who were better at reflective learning or reflexive learning also performed better in the training task, and the impact of reflective learning was stronger with more blocks. In addition, the marginally significant interactions indicated a trend that when reflexive learning was stronger, the positive impact of reflective learning on training performance was weaker than when reflexive learning was weaker, and vice versa (Figure 10 upper panels).

It is noteworthy that, as shown in Table 3, the variance of random effects of participants decreased from Model 2 to Model 3a/3b. This pattern shows that with the fixed effects of reflective/reflexive learning added in the model, there was a tradeoff between fixed effects and random effects. This might explain why the fixed effects of reflective/reflexive learning in Model 3a, 3b and Model 4 were significant, but the overall model fit did not improve compared to Model 2. Similarly, when comparing Model 3a/3b and Model 4, adding the interaction terms of reflective learning, reflexive learning, block, and mapping did not improve the overall model fit, but the variance of random effects of participants decreased, again showing a tradeoff between fixed effects and random effects.

Categorization task

Figure 8 shows the mean accuracy for the categorization task in O-P and O-S mappings in the AOL task. Similar to Experiment 1, performance in the O-S mapping~~s~~ was better than that in the O-P mapping for both trained and transfer words.

Mixed effects models were structured as in Experiment 1. Accuracy of AOL categorization was first modeled with fixed effects of word type (Model 1), and then mapping and the interaction terms of mapping were added (Model 2). Similar to the training task, the fixed effect and related interaction terms of reflective learning (Model 3a) and those of reflexive learning (Model 3b) were added. Lastly, both reflective and reflexive learning, as well as their interaction terms with mapping, block, and with each other were added (Model 4).

Model 1 included the fixed and random effects of word type. As in Experiment 1, word type was not a significant predictor. Model 2 showed that adding fixed and random effects of mapping significantly improved the model fit, $\chi^2(5) = 508.69$, $p < .001$. The significant fixed effect of mapping showed that performance in the O-S mapping was better than that in the O-P mapping, as shown in Figure 8, which was also consistent with the results in Experiment 1.

Figure 9 shows the scatterplots for the relationship between reflective/reflexive learning and performance in categorization task. Correlations between performance in categorization and reflective/reflexive learning were moderate, but the correlation between reflexive learning and categorization were weaker compared to reflective learning, which was the opposite to the findings in the training task.

Model 3a added fixed effects of reflective learning and Model 3b reflexive learning. As in the training task, adding reflective related fixed effects did not improve the model fit, $\chi^2(3) = 4.79$, $p = .19$, although reflective learning was a significant predictor in Model 3a. The results were consistent with the training task, showing a weak trend that participants with better reflective learning scores also performed better in the categorization task. In Model 3b, adding reflexive related fixed effects did not improve the model fit either, $\chi^2(3) = 2.43$, $p = .48$, which was also consistent with training task. Reflexive learning was not significant in the model either. The results showed that reflexive learning did not have an impact on categorization performance.

Similar to the model for training, Model 4 included all the fixed effects in Model 3a and 3b, as well as interaction between reflective and reflexive learning, and interaction between reflective learning, reflexive learning, and mapping. Model 4 improved the model fit of Model

3b but not 3a, $\chi^2(5) = 8.36$, $p = .14$ for 3a and $\chi^2(5) = 10.72$, $p = .057$ for 3b, indicating that reflective learning had stronger impact to categorization performance compared to reflexive learning. But when comparing Model 4 to Model 2 (which only included word type and mapping), Model 4 did not improve the model fit, $\chi^2(8) = 13.15$, $p = .11$, showing that the impact of reflective and reflexive learning on categorization performance was weak. In Model 4, similar to Model 3a and 3b, reflective learning was significant but reflexive learning was not. The interaction between reflective and reflexive learning was significant, which was consistent with the training task, indicating that when reflexive learning was stronger, the positive impact of reflective learning on training performance was weaker than when reflexive learning was weaker, and vice versa (Figure 10 lower panels).

Similar to the training task, Table 4 shows that the variance of random effects of participants decreased from Model 2 to Model 3a, exhibiting a tradeoff between fixed effects and random effects. The tradeoff was also shown when comparing Model 3a/3b and Model 4. Therefore, although reflective learning was significant in Model 3a and the interaction between reflective and reflexive learning was significant in Model 4, the model fit did not improve.

VSL and other learning tasks

Because of the robust relationship between performance in VSL and AOL tasks found in other studies and the complexity of the mechanisms of SL, Experiment 2 also examined the relationship between VSL and reflective/reflexive learning, and between VSL and AOL, to further understand SL and its relationship with learning to read. Before examining the relationship between VSL and other learning tasks, Pearson correlations between VSL and matrix reasoning (a measure of nonverbal IQ) and between VSL and working memory were measured, with $r = .14$ and $.15$, respectively, $ps > .05$, showing that VSL did not depend on these two general cognitive abilities, which was consistent with what Frost et al. (2013) found. The Pearson correlation between VSL and reflective learning, and between VSL and reflexive learning was .12 and .28 respectively, showing a stronger relationship between VSL and reflexive learning. This was supported by the multiple regression model using the two types of learning and their interaction to predict VSL scores, with the whole model explaining 13.1% of the variance ($F(3, 69) = 3.45$, $p < .05$), and reflexive learning was the only significant predictor ($\beta = .33$, $t = -2.71$, $p < .01$). The results were consistent with some of the previous findings (e.g., Perruchet & Pacton, 2006) that SL was more likely to be nonverbal, implicit learning.

The strength of correlations between performance of VSL and AOL tasks were moderate, indicating that better VSL scores were to some degree related to better AOL performance (VSL & AOL training collapsed over blocks and mappings, $r = .16$; VSL & AOL categorization collapsed over word types and mappings, $r = .17$). To examine the relationship between VSL and AOL performance, VSL scores were standardized and added to the mixed effects models after block terms and mapping for AOL training task, and after word type and mapping for categorization. For training, VSL was not a significant predictor. For categorization, VSL was marginally significant, Estimate = .26, $z = 1.80$, $p = .07$, indicating a trend that better VSL scores were related to better AOL categorization performance. Taking training and categorization together, the overall weak relationship between VSL and AOL performance was inconsistent with Rueckl et al. (2016) with the two tasks. The overall distribution of VSL scores in the current experiment shifted to the left compared to other studies using the same VSL task in Rueckl et al. (2016), indicating that the performance here was overall lower than in our other experiments, which might be the reason why the relationship between VSL and AOL was much weaker than previous experiments. It was surprising that Experiment 2 did not replicate Rueckl et al. (2016), but given the change in the task, there might be two reasons, which were also examined and would be further elaborated in discussion.

The relationship between reflective, reflexive learning and working memory

The relationship between working memory and reflective/reflexive learning performance was examined to understand previous conflicting evidence (DeCaro et al., 2008; Kalish et al., 2017). A positive correlation was found between working memory and reflective learning, $r = .34$, $p < .01$, and between working memory and reflexive learning, $r = .32$, $p < .01$. Moreover, the correlation between reflective and reflexive learning was also positive, $r = .31$, $p < .01$, indicating that participants who were good with one learning ability tended to be good in the other. The results were inconsistent with the DeCaro et al. (2008) findings that working memory was positively related to reflective learning but negatively related to reflexive learning, but supports the findings of Kalish et al. (2017) that working memory was positively related to both types of learning.

Discussion for Experiment 2

The results of Experiment 2 suggests that reflective learning was more engaged in doing AOL tasks than reflexive learning, which was shown by the evidence that when adding reflective and reflexive learning into the mixed effects models, there was a trend that both types of learning predicted AOL training performance, whereas only reflective learning but not reflexive learning predicted AOL categorization performance. This was consistent with results of Experiment 1 that AOL performance in CT condition was worse than Baseline condition. In addition, results from AOL tasks in Experiment 2 were consistent with findings in Experiment 1, showing that learning was better with more blocks in training, performance on trained and transfer words were similar in categorization, and overall learning in O-S mapping was always better than that in O-P mapping.

In Experiment 1, when reflective learning was disrupted by a concurrent task, performance in AOL tasks on O-S mapping was affected more than that on O-P mapping. This suggests two possibilities: 1. reflective learning was engaged more in O-S learning than in O-P learning, 2. O-P learning was already weak, so disturbing reflective learning cannot lower the performance much. In Experiment 2, no interaction between reflective learning and mapping was found, which does not support the first possibility. Therefore, results of this study did not provide evidence that reflective learning differentially supports O-P and O-S learning. Similarly, reflexive learning had no interaction with mapping in Experiment 2 either, therefore the results did not suggest that reflexive learning was engaged differently in O-P and O-S learning.

The interaction between reflective and reflexive learning was an important and interesting trend that found in Experiment 2, suggesting that with weaker reflexive learning, reflective learning was a stronger predictor to AOL performance, and similarly, with weaker reflective learning, reflexive learning was a stronger predictor to AOL performance. This is consistent with my speculation that in Experiment 1 disrupting reflexive learning with delayed feedback led to more engagement of reflective learning, which compensated the weakened reflexive learning. On the other hand, for the performance in CT condition, although performance of AOL tasks can also be compensated by reflexive learning, the compensation was limited since the engagement of reflexive learning was not very strong in the first place, so performance in CT condition was still worse than that in Baseline condition. This interaction provides evidence that both reflective and reflexive learning are underlie learning in the AOL paradigm, and also suggests a competition between reflective and reflexive learning.

Turning to a different aspect of the results, VSL was only a marginally significant predictor to AOL categorization, and not significant at all in predicting AOL training, which was against my expectation, and not consistent with findings in Rueckl et al. (2016). Comparing to Rueckl et al. (2016), the overall performance of VSL was poorer therefore individual differences were hard to find. This may be due to the change of order in which the AOL and VSL tasks were administered. In Rueckl et al. (2016) VSL was the first task; in contrast, because VSL was not the main focus of Experiment 2 and I wanted to avoid floor effects on the AOL tasks, in Experiment 2 the AOL protocol was administered first and VSL followed. Training and categorization in AOL took about 50 minutes, so fatigue may be a reason for the relatively poor VSL scores in Experiment 2. However, it should also be noted that the inter-trial interval in the AOL training task in Experiment 2 was longer than that used in the earlier studies. Although the inter-trial interval is unlikely to be related to the relatively poor VSL performance in my study, this difference may still affect performance in AOL training task therefore affect the relationship that between VSL and AOL.

To see which of the two possibility was true, a follow-up experiment was conducted with exactly the same VSL task, followed by two version of AOL tasks which was a between-subject condition, one with short ITI (500 ms as used before) and the other with long ITI (3500 ms as in Experiment 1 and 2). Mixed effects models showed that ITI did not predict performance in AOL tasks, but VSL was a significant predictor to AOL performance, both in training (Estimate = .16, $z = 4.35$, $p < .001$) and categorization (Estimate = .21, $z = 1.97$, $p < .05$). This follow-up experiment confirmed the relationship between VSL and AOL, and also showed that the reason VSL showed a different pattern in Experiment 2 was probably due to the change of the order of tasks.

Another secondary analysis was to examine the relationship between reflective/reflexive learning and working memory. DeCaro et al. (2008) found a positive relationship between working memory and reflective learning and a negative relationship between working memory and reflexive learning. Their interpretation was that individuals with better working memory capacity may tend to rely more on reflective learning, which is not the optimal strategy for the reflexive learning task. However, several recent studies suggested that working memory was positively related to both types of learning. For example, Lewandowsky, Yang, Newell and Kalish (2012) found that better working memory was related to better performance in both reflective and reflexive learning, and Kalish et al. (2017) suggested that working memory tends to aid performance in category learning tasks in general, regardless of the structure of categories. Because of the conflicting evidence, I did this secondary analysis to examine the relationship between working memory and reflective/reflexive learning, using the same learning tasks as in DeCaro et al. (2008). The working memory task I used was a modified auditory version of sentence span which was similar to what they used. What I found in Experiment 2 was consistent with Lewandowsky et al. (2012) and Kalish et al. (2017), suggesting that working memory was necessary for both reflective and reflexive learning.

General Discussion

Literature in theories and empirical studies of reading supports the idea that learning regularities is important for reading acquisition, but how people learn regularities is still unclear in the context of reading. The current study was inspired by the COVIS model (Ashby et al., 1998, Ashby & Waldron, 1999), which suggests that two independent systems are engaged in learning regularities. Individuals primarily rely on reflective learning when regularities are rule-based, whereas they primarily rely on reflexive learning when regularities are probabilistic or hard to detect through conscious awareness. This study was intended to answer whether learning to read involves reflective and/or reflexive learning. In addition, since O-P mapping has more regularities than O-S mapping and some theories suggest they are processed differentially, this study investigated whether O-P and O-S learning relies on different mechanisms.

In Experiment 1, two manipulations were used in the AOL training task – a delayed feedback and a concurrent task – to examine whether reflective learning and/or reflexive learning were engaged in learning regularities in the O-P and O-S mappings. The results showed that adding a concurrent task significantly weakened the performance of the AOL training and categorization, but delayed feedback did not affect the overall performance. In Experiment 2, reflective and reflexive learning were directly measured using visual category learning tasks. Reflective learning predicted performance on both AOL training and categorization, and reflexive learning predicted performance on AOL training but not categorization. Taken together, the findings of these two experiments suggest that both reflective and reflexive learning were used in AOL learning, but reflective learning was more important than reflexive learning.

An interaction between reflective and reflexive learning was significant in predicting both AOL learning and categorization. This may be explained by the competition between the two learning systems. Ashby, Queller and Berretty (1999) and Ashby et al. (2002) suggested that individuals assigned to the information-integration category learning task will be forced to switch to reflective learning when feedback is delayed because reflexive learning will be impaired. This switch of strategy has been confirmed by neuroimaging evidence (e.g., Arbel, Hong, Baker & Holroyd, 2017; Foerde & Shohamy, 2011) and the decision boundary models which compared participants' decision boundary to the ideal decision boundary for the category structure they are assigned to learn and to see how close they are (e.g., Ashby, 1992; Ashby &

Maddox, 2005; Ashby & Valentin, 2016; Ashby et al., 2011; Maddox & Ashby, 1993; Maddox, Ashby & Bohil, 2003; Smith et al., 2014; Smith, Jamani, Boomer & Church, 2018, but see Edmunds, Milton & Wills, 2018 for concern of the accuracy of this method). Therefore, when feedback was delayed in the AOL training, it was possible that reflexive learning was disrupted and reflective learning was more engaged. Since reflexive learning was only weakly engaged in the AOL tasks as suggested by Experiment 2, the loss was easily compensated for by reflective learning, so that the overall performance was not affected. In addition, there was an unexpected nonsignificant trend in the DF condition, in which O-P learning was better than in the Baseline condition. If this trend is true, it might be because the three-seconds delay may have helped participants be more focused while waiting and may have helped them figure out the regularities, which would involve reflective learning but not reflexive learning. Therefore, although the delayed feedback weakened reflexive learning, reflective learning may have been strengthened. In this situation, it is not unreasonable to assume that this would occur.

Connectionist models suggest that a single process underlies the computation of statistical information for both O-P and O-S learning. Alternatively, in real world reading, since the correspondences between words' form and meaning are more arbitrary compared to the correspondences between word form and pronunciation, some theories suggest learning the two kinds of mappings can be different (e.g., Ullman & Pierpont, 2005). Although Experiment 1 showed an interaction between condition and mapping, no interaction between reflective or reflexive learning and mapping was found in Experiment 2. Therefore, the interaction found in Experiment 1was probably due to different performance in the O-P and O-S learning in the task; i.e., the stronger impact of the concurrent task on O-S leaning was because O-P learning was weak. Combined together, the current study did not support that reflective and reflexive learning were used differentially in learning O-P and O-S mappings.

In addition, working memory was found to positively correlated with both reflective and reflexive learning, which supports Lewandowsky et al. (2012) and Kalish et al. (2017), suggesting that working memory was necessary for both reflective and reflexive learning. The relationship between reflective/reflexive learning and VSL performance was also examined to understand more about the mechanism of VSL. However, in the current study the overall performance on VSL task was poorer than the previous findings in Rueckl et al. (2016), making it more difficult to discover reliable individual differences in VSL performance. A followup analysis suggested that it was probably due to the order of the tasks: VSL task was conducted after the AOL tasks which took about fifty minutes, and more individuals may have been around the lower end of the distribution of VSL performance due to fatigue. Even with this atypical VSL performance, reflexive learning was still a good predictor to VSL. Reflexive learning and implicit learning are not exactly the same thing, but they are closely related, because, like implicit learning, reflexive learning is not affected by conscious awareness. This is not strong evidence, but it still supports earlier findings about the close relationship between SL and implicit learning (e.g., Perruchet & Pacton, 2006).

Although this study provided insightful results, there are some limitations in its methodology. First, as mentioned earlier, the delayed-feedback manipulation in Experiment 1 may encourage reflective learning, which may be the reason why delayed-feedback did not affect the overall performance in the AOL tasks. Second, the ceiling in reflective learning reflects a potential problem for reliability. Third, there are potential problems in the validity of the visual category learning tasks. There is no standard task to measure reflective and reflexive learning, so

I borrowed the tasks from DeCaro et al. (2008). However, the particular reflective and reflexive learning tasks cannot guarantee that people use one type of learning over the other. Especially in reflexive learning task, although the optimal strategy is to integrate information from three dimensions to categorize the items, it is still possible to get 75% correct if participants only use one dimension. For example, in the reflexive learning task shown in Figure 3, if participants adopted a simpler, one-dimension-based strategy which considered all the items with green symbols as category A and all the items with red symbol as category B, then they would be correct three out of four times. To examine what strategies participants used with this task, DeCaro, Carlson, Thomas and Beilock (2009) examined the extent to which participants' responses matched four different possible strategies during reflexive learning and compared the responses on each learning trial with the predicted responses from each possible strategy. They concluded that participants did not rely on a simple strategy, but explored the optimal strategy with more trials. Although their findings suggest that this visual categorization task was adequate to measure reflexive learning, it can still be a potential problem since we cannot be sure our participants followed the same strategy.

The findings of this study suggest that reflective learning was used more to pick up on the regularities in AOL learning as compared to reflexive learning. However, we still need to be cautious when generalizing the findings to learning to read in the real world. The regularities built into the AOL tasks were highly reliable with no exceptions. However, quasi-regularities are much more common in natural languages, which probably encourages the use of reflexive learning since it is mostly used to deal with complex or probabilistic regularities. Moreover, writing systems with different orthographic depth are likely to encourage different strategies in using reflective and reflexive learning. Shallow languages with more reliable regularities like

Spanish and German tend to encourage learning the regularities explicitly through reflective learning, whereas deep languages like English and Hebrew that have more irregular cases tend to encourage learning those more probabilistic regularities through reflexive learning.

In natural languages, O-P mappings have more regularities than O-S mappings. This is different from my AOL tasks in which regularities in O-P and O-S mappings were symmetric. Learning arbitrary correspondences in the O-S mapping may involve learning systems other than reflective and reflexive learning since these two discussed in the current study are about learning regularities. The DP model proposed by Ullman and others (e.g., Ullman, 2001, 2004; Ullman & Pierpont, 2005) suggests that declarative learning underlies learning idiosyncratic forms and more arbitrary correspondences. So, when generalizing the question to learning to read in a natural situation, reflective and reflexive learning should not be the only systems we need to consider.

The time course of learning is another big difference between this study and a natural learning situation. My participants only had thirty minutes of training in the AOL task; learning to read usually takes years of efforts. Since reflexive learning needs repetition and occurs over time, its importance may be more significant over longer periods of time. Also, some evidence suggests that the age of the learners may also affect how they learn with different learning systems. For example, in the frame of DP model, Ullman (2001) suggests that children are more likely to use the procedural learning system to learn their native language compared to adults learning a second language. Therefore, although there is no direct evidence about how they use reflective and reflexive learning systems, we need to consider the developmental trajectory when theorizing how individuals use different learning systems.

Findings in this study also have implications in teaching children to read. Educators have disagreed about how the regularities, especially the O-P correspondences should be taught in alphabetic languages. When they enter school, children usually have already possessed substantial capability of speaking in their language, but have only little knowledge of reading and writing. The question is that in order to link children's knowledge of spoken language to written language, whether teachers should adopt a systematic phonics approach in which explicit instructions about the systematicity in O-P mapping are given to children, or the whole-language approach in which the correspondences are not taught explicitly and are incidentally in context (e.g., Ehri, Nunes, Stahl & Willows, 2001; Rayner, Foorman, Perfetti, Pesetsky & Seidenberg, 2001). Some researchers have advocated the former approach because evidence suggests that it is more effective than a non-systematic approach. For example, Ehri et al. (2001) conducted a meta-analysis to evaluate the effectiveness of a systematic phonics approach compared to a non-systematic approach by examining 66 treatment-control comparisons from 38 experiments. They confirmed that the systematic approach was more effective, particularly when they were used earlier (kindergarten or first-grade) rather than later (after first-grade). Rayner et al. (2001) summarized the research in teaching reading in the classroom and concluded that the phonetics approach was more effective in helping children master the O-P mapping compared to the whole-language approach. In addition, the whole-language approach can be used to supplement the phonetics approach for they can make the O-P correspondences taught to children more meaningful by providing context. In sum, although some studies suggest that phonics is boring for children and may harm their interest in reading (e.g., Anderson, Wilson & Fielding, 1988; McArthur & Castles, 2017), most evidence supports that the phonetics approach helps children in word reading, spelling, and text comprehension (Castles, Rastle & Nation, 2018; Ehri et al.

2001). By showing the importance of reflective learning in learning O-P and O-S regularities, the current study provides further evidence supporting the phonetics approach. Since the participants in the current study were college students, it also suggests that the phonetics approach may improve learning to read not only for children but for adults learning a second language.

Taken together, the main findings of the current study suggest that in early stages of learning a second language, learning O-P and O-S correspondences in the AOL task~~s~~ mostly relies on reflective learning. Reflexive learning may also be engaged, but its impact is weaker compared to reflective learning. The two learning systems compete with each other, which means that when one system gets disrupted, the other will be enhanced. No evidence in this study suggests that learning regularities in O-P and O-S mappings rely on different mechanisms.

References

Aaron, P. G., Joshi, R. M., Boulware-Gooden, R., & Bentum, K. (2008). Diagnosis and treatment of reading disabilities based on the component model of reading: An alternative to the discrepancy model of learning disabilities. *Journal of Learning Disabilities, 41,* 67–84.

Alba, D., & Okanoya, K. (2008). Statistical segmentation of tone sequences activates the left inferior frontal cortex: A near-infrared spectroscopy study. *Neuropsychologia, 46*(11), 2787–2795.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*(4), 369–406.

Anderson, R. C., Wilson, P. T., & Fielding, L. G. (1988). Growth in reading and how children spend their time outside of school. *Reading Research Quarterly, 23,* 285–303.

Andrews, S. (1982). Phonological recoding: Is the regularity effect consistent? *Memory & Cognition, 10,* 565-575.

Araújo, S., Reis, A., Petersson, K. M., & Faísca, L. (2015). Rapid automatized naming and reading performance: A meta-analysis. *Journal of Educational Psychology, 107*(3), 868-883.

Arbel, Y., Hong, L., Baker, T. E., & Holroyd, C. B. (2017). It's all about timing: an electrophysiological examination of feedback-based learning with immediate and delayed feedback. *Neuropsychologia, 99,* 179-186.

Arciuli, J. (2018). Reading as statistical learning. *Language, Speech, and Hearing Services in Schools, 49*(3S), 634–643.

Arciuli, J., & Simpson, I. C. (2012). Statistical learning is related to reading ability in children and adults. *Cognitive Science, 36*(2), 286–304.

Armstrong, B. C., Dumay, N., Kim, W., & Pitt, M. A. (2017). Generalization from newly learned words reveals structural properties of the human reading system. *Journal of Experimental Psychology: General, 146*(2), 227-249.

Armstrong, B. C., Frost, R., & Christiansen, M. H. (2017). The long road of statistical learning research: past, present and future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*, 20160047.

Ashby, F. G. (1992). Multidimensional models of categorization. In F. G. Ashby (Ed.), *Multidimensional Models of Perception and Cognition* (pp. 449 – 484). Hillsdale, NJ: Erlbaum.

Ashby, F. G., & Crossley, M. J. (2010). Interactions between declarative and procedural-learning categorization systems. *Neurobiology of Learning and Memory*, *94*, 1–12.

Ashby, F. G., Isen, A. M., Turken, A. U. (1999a). A neuropsychological theory of positive affect and its influence on cognition. *Psychological Review, 106,* 529–550.

Ashby, F. G., Maddox, W. T., Bohil, C. J. (2002). Observational versus feedback training in rule-based and information-integration category learning. *Memory & Cognition*, *30,* 666–677.

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology*, *56,* 149-178.

Ashby, F. G., Maddox, W. T., & Bohil, C. J. (2002). Observational versus feedback training in rule-based and information-integration category learning. *Memory & Cognition, 30*, 666-667.

Ashby, F. G., Queller, S., & Berretty, P. T. (1999). On the dominance of unidimensional rules in unsupervised categorization. *Perception & Psychophysics, 61*, 1178-1199.

Ashby, F. G., & Valentin, V. V. (2016) Multiple systems of perceptual category learning: Theory and cognitive tests. In H. Cohen & C. Lefebvre (Eds.), *Hand-book of Categorization in Cognitive Science* (pp. 547–572) (2nd ed.). New York, NY: Elsevier.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language, 59(4),* 390-412.

Baldwin, D., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition, 106,* 1382–1407.

Batterink, L. J., Reber, P. J., Neville, H. J., & Paller, K. A. (2015). Implicit and explicit contributions to statistical learning. *Journal of memory and language, 83,* 62-78.

Berry, D., & Dienes, Z.P. (1993). *Implicit learning: Theoretical and empirical issues.* Hove, East Sussex: Psychology Press.

Bertels, J., Franco, A., & Destrebecqz, A. (2012). How implicit is visual statistical learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38(5),* 1425-1431.

Bradley, L., & Bryant, P.E. (1983). Categorizing sounds and learning to read – a causal connection. *Nature, 301,* 419–421.

Breitenstein, C., Jansen, A., Deppe, M., Foerster, A. F., Sommer, J., Wolbers, T., & Knecht, S. (2005). Hippocampus activity differentiates good from poor learners of a novel lexicon. *Neuroimage, 25(3),* 958-968.

Caravolas M, Lervåg A, Mousikou P, Efrim C, Litavský M, Onochie-Quintanilla E, et al. (2012). Common patterns of prediction of literacy development in different alphabetic orthographies. *Psychological Science, 23,* 678–686.

Carlisle, J. F. (2000). Awareness of the structure and meaning of morphologically complex words: Impact on reading. *Reading and Writing, 12,* 169-190.

Cassar, M., & Treiman, R. (1997). The beginnings of orthographic knowledge: Children's knowledge of double letters in words. *Journal of Educational Psychology, 89,* 631-644.

Castles, A., & Coltheart, M. (1993). Varieties of developmental dyslexia. *Cognition, 47*, 149–180.

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest, 19(1),* 5-51.

Coltheart, M. (2006). Dual route and connectionist models of reading: An overview. *London Review of Education, 4(1),* 5-17.

Coltheart, M. (2015). What kinds of things cause children's reading difficulties? *Australian Journal of Learning Difficulties, 20,* 103-112.

Coltheart, M., Curtis, B., Atkins, P. & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review, 100,* 589–608.

Coltheart, M., Rastle, K., Perry, C., Langdon, R. & Ziegler, J. C. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*, 204–256.

Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between modalities: Pitting abstract against stimulus-specific representations. *Psychological Science, 17(10)*, 905-912.

Conway, C. M., Arciuli, J., Lum, J. A., & Ullman, M. T. (2019). Seeing problems that may not exist: A reply to West et al.'s (2018) questioning of the procedural deficit hypothesis. *Developmental Science*, e12814.

Daltrozzo, J., Emerson, S. N., Deocampo, J., Singh, S., Freggens, M., Branum-Martin, L., & Conway, C. M. (2017). Visual statistical learning is related to natural language ability in adults: an ERP study. *Brain and language, 166,* 40-51.

Daneman, M, & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450–466.

Deacon, S. H., Conrad, N., & Pacton, S. (2008). A statistical learning perspective on children's learning about graphotactic and morphological regularities in spelling. *Canadian Psychology/Psychologie Canadienne, 49(2),* 118.

Deacon, S. H., & Leung, D. (2013). Testing the statistical learning of spelling patterns by manipulating semantic and orthographic frequency. *Applied Psycholinguistics, 34,* 1093-1108.

DeCaro, M. S., Carlson, K. D., Thomas, R. D., & Beilock, S. L. (2009). When and how less is more: Reply to Tharp and Pickering. *Cognition, 111(3),* 415-421.

DeCaro, M. S., Thomas, R. D., Albert, N. B., & Beilock, S. L. (2011). Choking under pressure: Multiple routes to skill failure. *Journal of Experimental Psychology: General, 140(3),* 390.

Dunn, J. C., Newell, B. R., & Kalish, M. L. (2012). The effect of feedback delay and feedback type on perceptual category learning: the limits of multiple systems. Journal of Experimental Psychology: Learning, Memory, and Cognition, 38(4), 840-859.

Edmunds, C. E., Milton, F., & Wills, A. J. (2018). Due Process in Dual Process: Model‐Recovery Simulations of Decision‐Bound Strategy Analysis in Category Learning. *Cognitive Science, 42,* 833-860.

Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. *Review of Educational Research, 71(3),* 393-447.

Eichenbaum, H., & Cohen, N. J. (2001). *From Conditioning to Conscious Recollection: Memory Systems of the Brain.* New York, NY: Oxford University Press.

Elman, J. L., Bates, E. A., & Johnson, M. H. (1998). *Rethinking innateness: A connectionist perspective on development* (Vol. 10). MIT press.

Erickson, M. A. (2008). Executive attention and task switching in category learning: evidence for stimulus-dependent representation. *Memory & Cognition, 36*, 749–761.

Fiser, J., & Aslin, R. N. (2002a). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 458–467.

Fiser, J., & Aslin, R. N. (2002b). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences, 99,* 15822–15826.

Folia, V., Uddén, J., Forkstam, C., Ingvar, M., Hagoort, P., & Petersson, K. M. (2008). Implicit learning and dyslexia. *Annals of the New York Academy of Sciences, 1145(1),* 132–150.

Franco, A., Cleeremans, A., & Destrebecqz, A. (2011). Statistical learning of two artificial languages presented successively: How conscious? *Frontiers in Psychology, 2,* 229.

Frank, M. J., O'Reilly, R. C., & Curran, T. (2006). When memory fails, intuition reigns: Midazolam enhances implicit inference in humans. *Psychological Science, 17,* 700–707.

Frost, R., Siegelman, N., Narkiss, A., & Afek, L. (2013). What predicts successful literacy acquisition in a second language? *Psychological Science, 24(7),* 1243–1252.

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of experimental psychology: Human perception and performance, 5*, 674-691.

Gombert, J. (2003). Implicit and explicit learning to read: implication as for subtypes of dyslexia. *Current Psychology Letters, 10(1),* 25–40.

Gómez, R. L., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy, 7,* 183–206.

Goodroe, S. C., Starnes, J. M., & Brown, T. I. (2018). The complex nature of hippocampal-striatal interactions in spatial navigation. *Frontiers in Human Neuroscience, 12,* 250.

Harm, M. and Seidenberg, M. S. 1999. Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review, 106*, 491–528.

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review, 111(3),* 662–720.

Hayes, H., Treiman, R., & Kessler, B. (2006). Children use vowels to help them spell consonants. *Journal of Experimental Child Psychology, 94(1),* 27-42.

Henderson, L. M., & Warmington, M. (2017). A sequence learning impairment in dyslexia? It depends on the task. *Research in Developmental Disabilities, 60,* 198–210.

Howard, D. V., Howard, J. H., Jr., Japikse, K., DiYanni, C., Thompson, A., & Somberg, R. (2004). Implicit sequence learning: Effects of level of structure, adult age, and extended practice. *Psychology and Aging, 19(1),* 79–92

Howard, J. H., Howard, D. V., Japikse, K. C., & Eden, G. F. (2006). Dyslexics are impaired on implicit higher-order sequence learning, but not on implicit spatial context learning. *Neuropsychologia, 44,* 1131–1144.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing, 2,* 127-160.

Hulme, C., Bowyer-Crane, C., Carroll, J. M., Duff, F. J., & Snowling, M. J. (2012). The causal role of phoneme awareness and letter-sound knowledge in learning to read: Combining intervention studies with mediation analyses. *Psychological Science, 23(6),* 572-577.

Hulme, C., & Snowling, M. J. (2013). Learning to read: What we know and what we need to understand better. *Child Development Perspectives, 7(1),* 1-5.

Inácio, F., Faísca, L., Forkstam, C., Araújo, S., Bramão, I., Reis, A., & Petersson, K. M. (2018). Implicit sequence learning is preserved in dyslexic children. *Annals of Dyslexia, 68(1),* 1-14.

Jimenez-Fernandez, G., Vaquero, J. M., Jimenez, L., & Defior, S. (2011). Dyslexic children show deficits in implicit sequence learning, but not in explicit sequence learning or contextual cueing. *Annals of Dyslexia, 61,* 85–110.

Johansson, T. (2009). Strengthening the case for stimulus-specificity in artificial grammar learning: no evidence for abstract representations with extended exposure. *Experimental Psychology, 56(3),* 188-197.

Joshi, R. M., & Aaron, P. G. (2000). The component model of reading: Simple view of reading made a little more complex. *Reading Psychology, 21,* 85–97.

Kahta, S., & Schiff, R. (2016). Implicit learning deficits among adults with developmental dyslexia. *Annals of Dyslexia, 66(2),* 235-250.

Kalish, M. L., Newell, B. R., & Dunn, J. C. (2017). More is generally better: Higher working memory capacity does not impair perceptual category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43(4),* 503-514.

Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (2012). *Principles of Neural Science (5th ed.).* New York, NY/London: McGraw-Hill.

Karuza, E. A., Newport, E. L., Aslin, R. N., Starling, S. J., Tivarus, M. E., & Bavelier, D. (2013). The neural correlates of statistical learning in a word segmentation task: An fMRI study. *Brain and Language, 127(1),* 46–54.

Kelly, S. W., Griffiths, S., & Frith, U. (2002). Evidence for implicit sequence learning in dyslexia. *Dyslexia, 8,* 43–52

Kim, R., Seitz, A., Feenstra, H., & Shams, L. (2009). Testing assumptions of statistical learning: is it long-term and implicit?. *Neuroscience Letters, 461(2),* 145-149.

Kirby, J. R., Georgiou, G. K., Martinussen, R., & Parrila, R. (2010). Naming speed and reading: From prediction to instruction. *Reading Research Quarterly, 45,* 341–362.

Kóbor, A., Takács,Á., Kardos, Z., Janacsek, K., Horváth, K., Csépe, V., & Nemeth, D. (2018). ERPs differentiate the sensitivity to statistical probabilities and the learning of sequential structures during procedural learning. *Biological Psychology, 135,* 180–193.

Kuperman, V., & Van Dyke, J. A. (2011). Effects of individual differences in verbal skills on eye-movement patterns during sentence reading. *Journal of memory and language, 65(1),* 42-73.

Lum, J. A., Ullman, M. T., & Conti-Ramsden, G. (2013). Procedural learning is impaired in dyslexia: Evidence from a meta-analysis of serial reaction time studies. *Research in Developmental Disabilities, 34(10),* 3460-3476.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101(4),* 676-703.

Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception and Psychophysics, 53,* 49-70.

Maddox, W. T., Ashby, F. G., & Bohil, C. J. (2003). Delayed feedback effects on rule-based and information-integration category learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 29(4),* 650–662.

Maddox, W. T., Ashby, F. G., Ing, A. D., & Pickering, A. D. (2004). Disrupting feedback processing interferes with rule-based but not information-integration category learning. *Memory & Cognition, 32,* 582-591.

Maddox, W. T., Filoteo, J. V., Hejl, K. D., Ing, A. D. (2004). Category number impacts rule-based but not information-integration category learning: further evidence for dissociable category learning systems. *Journal of Experimental Psychology: Learning, Memory & Cognition, 30,* 227–235.

Maddox, W. T., & Ing, A. D. (2005). Delayed feedback disrupts the procedural-learning system but not the hypothesis testing system in perceptual category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31(1),* 100–107.

Maddox, W. T., Love, B. C., Glass, B. D., & Filoteo, J. V. (2008). When more is less: Feedback effects in perceptual category learning. *Cognition, 108(2),* 578–589.

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science, 11,* 122–134.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82,* B101–B111.

McArthur, G., & Castles, A. (2017). Helping children with reading difficulties: Some things we have learned so far. *Npj Science of Learning, 2(1),* Article 7.

Melby-Lervåg, M., Lyster, S. A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin, 138*, 322-352.

Menghini, D., Hagberg, G. E., Caltagirone, C., Petrosini, L., & Vicari, S. (2006). Implicit learning deficits in dyslexic adults: An fMRI study. *NeuroImage, 33(4),* 1218-1226.

Misyak, J. B., Christiansen, M. H. (2012). Statistical learning and language: an individual differences study. *Language Learning, 62,* 302–331.

Moody, T. D., Bookheimer, S. Y., Vanek, Z., & Knowlton, B. J. (2004). An implicit learning task activates medial temporal lobe in patients with Parkinson's disease. *Behavioral Neuroscience, 118(2),* 438-442.

Nemeth, D., Janacsek, K., & Fiser, J. (2013). Age-dependent and coordinated shift in performance between implicit and explicit skill learning. *Frontiers in Computational Neuroscience. 7,* 147.

Nicolson, R. I., & Fawcett, A. J. (1990). Automaticity: A new framework for dyslexia research?. *Cognition, 35(2),* 159-182.

Nicolson, R. I., & Fawcett, A. J. (1994). Comparison of deficits in cognitive and motor skills among children with dyslexia. *Annals of dyslexia, 44(1),* 147-164.

Nicolson, R.I., & Fawcett, A.J. (2007). Procedural learning difficulties: Reuniting the developmental disorders? *Trends in Neurosciences, 30*, 135–141.

Nicolson, R.I., & Fawcett, A.J. (2010). *Dyslexia, Learning, and the Brain.* Cambridge, MA: MIT Press.

Nicolson, R.I., & Fawcett, A.J. (2011). Dyslexia, dysgraphia, procedural learning and the cerebellum. *Cortex, 47*, 117–127.

Nicolson, R. I., Fawcett, A. J., Brookes, R. L., & Needle, J. (2010). Procedural learning and dyslexia. *Dyslexia, 16(3)*, 194-212.

Nicolson, R. I., Fawcett, A. J., & Dean, P. (2001). Developmental dyslexia: The cerebellar deficit hypothesis. *Trends in Neurosciences, 24,* 508–511.

Nigro, L., Jiménez-Fernández, G., Simpson, I. C., & Defior, S. (2016). Implicit learning of non-linguistic and linguistic regularities in children with dyslexia. *Annals of dyslexia, 66(2),* 202-218.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: evidence from performance measures. *Cognitive Psychology, 19(1),* 1–32.

Nomura, E. M., Maddox, W. T., Filoteo, J. V., Ing, A. D., Gitelman, D. R., Parrish, T. B., ... & Reber, P. J. (2007). Neural correlates of rule-based and information-integration visual category learning. *Cerebral Cortex, 17,* 37-43.

Norton, E. S., & Wolf, M. (2012). Rapid automatized naming (RAN) and reading fluency: Implications for understanding and treatment of reading disabilities. *Annual Review of Psychology, 63,* 427-452.

Packard, M. G., & Knowlton, B. J. (2002). Learning and memory functions of the basal ganglia. *Annual Review of Neuroscience, 25,* 563–593.

Pacton, S., Fayol, M., & Perruchet, P. (2005). Children's implicit learning of graphotactic and morphological regularities. *Child Development, 76,* 324–339.

Pacton, S., Perruchet, P., Fayol, M., & Cleeremans, A. (2001). Implicit learning out of the lab: The case of orthographic regularities. *Journal of Experimental Psychology: General, 130,* 401–426.

Pavlidou, E. V., Kelly, M. L., & Williams, J. M. (2010). Do children with developmental dyslexia have impairments in implicit learning? *Dyslexia, 16,* 143–161.

Pavlidou, E. V., & Williams, J. M. (2010). Developmental dyslexia and implicit learning: evidence from an AGL transfer study. *Procedia-Social and Behavioral Sciences, 2(2),* 3289-3296.

Pennington, B. F. (2006). From single to multiple deficit models of developmental disorders. *Cognition, 101(2),* 385–413.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences, 10,* 233-238.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103(1),* 56–115.

Poldrack, R. A., Clark, J., Pare-Blagoev, E. J., Shohamy, D., Moyano, J. C., Myers, C., & Gluck, M. A. (2001). Interactive memory systems in the human brain. *Nature, 414,* 546-550.

Poldrack, R. A., Prabhakaran, V., Seger, C. A., & Gabrieli, J. D. (1999). Striatal activation during acquisition of a cognitive skill. *Neuropsychology, 13(4),* 564-574.

Qi, Z., Sanchez Araujo, Y., Georgan, W. C., Gabrieli, J. D., & Arciuli, J. (2019). Hearing matters more than seeing: A cross-modality study of statistical learning and reading ability. *Scientific Studies of Reading, 23(1),* 101-115.

Quercia, P., Feiss, L., & Michel, C. (2013). Developmental dyslexia and vision. *Clinical Ophthalmology, 7,* 869–881.

Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest, 2(2),* 31-74.

Rauschecker, A. M., Pringle, A., & Watkins, K. E. (2008). Changes in neural activity associated with learning to articulate novel auditory pseudowords by covert repetition. *Human brain mapping, 29(11),* 1231-1242.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General, 118(3),* 219-235.

Reber, P. J. (2008). Cognitive neuroscience of declarative and nondeclarative memory. *Advances in Psychology, 139,* 113-123.

Redington, M., & Chater, N. (1996). Transfer in artificial grammar learning: A reevaluation. *Journal of experimental psychology: General, 125(2),* 123-138.

Ripollés, P., Marco-Pallarés, J., Hielscher, U., Mestres-Missé, A., Tempelmann, C., Heinze, H. J., ... & Noesselt, T. (2014). The role of reward in word learning and its implications for language acquisition. *Current Biology, 24(21),* 2606-2611.

Romberg, A. R., & Saffran, J. R. (2013). All together now: concurrent learning of multiple structures in an artificial language. *Cognition Science, 37,* 1290–1320.

Rueckl, J. G. (2016). Toward a theory of variation in the organization of the word reading system. *Scientific Studies of Reading, 20(1),* 86-97.

Rueckl, J., Li, T., Brown, J. & Zhao, J. (2016, November). The Relationship Between Visual Statistical Learning and Learning in an Artificial Orthography Task. Poster session presented at the Annual Meeting of the Psychonomic Society, Boston, MA.

Rueckl, J. G., & Raveh, M. (1999). The influence of morphological regularities on the dynamics of a connectionist network. *Brain and Language, 68(1–2),* 110–117.

Rüsseler, J., Gerth, I., & Münte, T. F. (2006). Implicit learning is intact in adult developmental dyslexic readers: Evidence from the serial reaction time task and artificial grammar learning. *Journal of Clinical and Experimental Neuropsychology, 28,* 808–827.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274,* 1926–1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70,* 27–52.

Sawi, O. M., & Rueckl, J. (2019). Reading and the neurocognitive bases of statistical learning. *Scientific Studies of Reading, 23,* 8-23.

Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13(3),* 501-518.

Schapiro, A., & Turk-Browne, N. (2015). Statistical learning. *Brain Mapping, 3,* 501–506.

Schmalz, X., Moll, K., Mulatti, C., & Schulte-Körne, G. (2018). Is statistical learning ability related to reading ability, and if so, why? *Scientific Studies of Reading, 23,* 1-7.

Seger, C. A., & Cincotta, C. M. (2005). Dynamics of frontal, striatal, and hippocampal systems during rule learning. *Cerebral Cortex, 16(11),* 1546-1555.

Seger, C. A., & Miller, E. K. (2010). Category learning in the brain. *Annual Review of Neuroscience, 33,* 203-219.

Seidenberg, M. S. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science, 275(5306),* 1599-1603.

Shanks, D.R. (2005). Implicit learning. In K. Lamberts & R.L. Goldstone (Eds.), *The Handbook of Cognition* (pp. 202–220). London: Sage.

Siegelman, N., Bogaerts, L., Christiansen, M. H., & Frost, R. (2017). Towards a theory of individual differences in statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences, 372(1711),* 20160059.

Siegelman, N, & Frost, R. (2015). Statistical learning as an individual ability: theoretical perspectives and empirical evidence. *Journal of Memory and Language, 81,* 105–120.

Smith, J. D., Boomer, J., Zakrzewski, A. C., Roeder, J. L., Church, B. A., & Ashby, F. G. (2014). Deferred feedback sharply dissociates implicit and explicit category learning. *Psychological Science, 25(2),* 447-457.

Smith, J. D., Jamani, S., Boomer, J., & Church, B. A. (2018). One-back reinforcement dissociates implicit-procedural and explicit-declarative category learning. *Memory & cognition, 46(2),* 261-273.

Snowling, M. J. (1987). *Dyslexia: A cognitive developmental perspective.* Oxford: Blackwell.

Snowling, M. (2000). *Dyslexia (2nd ed.).* Oxford, UK: Blackwell Publishers Inc.

Snowling, M. J. (2005). Literacy outcomes for children with oral language impairments: Developmental interactions between language skills and learning to read. In H. W. Catts, & A. G. Kamhi (Eds.), *The Connections Between Language and Reading Disabilities* (pp. 48-65). London & New Jersey: Lawrence Erlbaum Associates.

Spencer, M., Kaschak, M. P., Jones, J. L., & Lonigan, C. J. (2015). Statistical learning is related to early literacy-related skills. *Reading and Writing, 28(4),* 467–490.

Squire, L. R. (1992). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. *Journal of Cognitive Neuroscience, 4(3)*, 232–243.

Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory, 82*, 171–177.

Squire, L. R., & Knowlton, B. J. (2000). The medial temporal lobe, the hippocampus, and the memory systems of the brain. In M. S. Gazzaniga (Ed), *The New Cognitive Neurosciences* (pp. 765-780). Cambridge, MA: MIT Press.

Staels, E., & Van den Broeck, W. (2015). No solid empirical evidence for the solid (serial order learning impairment) hypothesis of dyslexia. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41,* 650-669.

Stanovich, K. E. (1988). Explaining the differences between the dyslexic and the garden-variety poor reader: The phonological-core variable-difference model. *Journal of Learning Disabilities, 21,* 590–604.

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior, 14(6),* 638–647.

Treiman, R., & Kessler, B. (2006). Spelling as statistical learning: Using consonantal context to spell vowels. *Journal of Educational Psychology, 98,* 642-652.

Tunney, R. J., & Altmann, G. (1999). The transfer effect in artificial grammar learning: Reappraising the evidence on the transfer of sequential dependencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25(5),* 1322-1333.

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience, 21(10),* 1934-1945.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*, 127–154.

Ullman, M.T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, *92*, 231–270.

Ullman, M. T. (2005). A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. In C. Sanz (Ed.), *Mind and Context in Adult Second Language Acquisition: Methods, Theory and Practice* (pp. 141–178). Washington, DC: Georgetown University Press.

Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex, 41(3),* 399-433.

Vakil, E., Lowe, M., & Goldfus, C. (2015). Performance of Children With Developmental Dyslexia on Two Skill Learning Tasks—Serial Reaction Time and Tower of Hanoi Puzzle: A Test of the Specific Procedural Learning Difficulties Theory. *Journal of Learning Disabilities, 48(5),* 471-481.

van der Kleij, S. W., Groen, M. A., Segers, E., & Verhoeven, L. (2018). Sequential implicit learning ability predicts growth in reading skills in typical readers and children with dyslexia. *Scientific Studies of Reading, 23(1),* 77–88.

Vellutino, F. R. (1979). *Dyslexia: Research and Theory.* Cambridge, MA: MIT Press.

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of Child Psychology and Psychiatry, 45,* 2–40.

Vicari, S., Finzi, A., Menghini, D., Marotta, L., Baldi, S., & Petrosini, L. (2005). Do children with developmental dyslexia have an implicit learning deficit? *Journal of Neurology, Neurosurgery, and Psychiatry, 76,* 1392–1397.

Vicari, S., Marotta, L., Menghiai, D., Molinari, M., & Petrosini, L. (2003). Implicit learning deficit in children with developmental dyslexia. *Neuropsychologia, 41,* 108–114.

Waber, D. P., Marcus, D. J., Forbes, P. W., Bellinger, D. C., Weiler, M. D., Sorensen, L. G., & Curran, T. (2003). Motor sequence learning and reading ability: Is poor reading associates with sequencing deficits? *Journal of Experimental Child Psychology, 84,* 338–354.

Waldron, E. M., & Ashby, F. G. (2001). The effects of concurrent task interference on category learning. *Psychonomic Bulletin & Review, 8,* 168–176.

West, G., Vadillo, M. A., Shanks, D. R., & Hulme, C. (2018). The procedural learning deficit hypothesis of language learning disorders: we see some problems. *Developmental science, 21(2),* e12552.

Wilkinson, G. S., & Robertson, G. J. (2006). *Wide range achievement test – Fourth edition: Professional manual*. Lutz, FL: Psychological Assessment Resources.

Wimmer, H., Mayringer, H., & Landerl, K. (2000). The double-deficit hypothesis and difficulties in learning to read a regular orthography. *Journal of Educational Psychology, 92,* 668–680.

Wu, D. H., Y Shih, E. H., Frost, R., Lee, J.-R., Lee, C.-Y., Tsai, J.-L., . . . Tzeng, O. J.-L. (2012, November). Statistical learning of implicit regularity correlates with acquisition of Chinese literacy as a second language. Paper presented at the annual meeting of the Psychonomic Society, Minneapolis, MN.

Zeithamova, D., & Maddox, W. T. (2007). The role of visuospatial and verbal working memory in perceptual category learning. *Memory & Cognition*, *35,* 1380-1398.

Zhao, J., Li, T., Elliott, M. A., & Rueckl, J. G. (2018). Statistical and Cooperative Learning in Reading: An Artificial Orthography Learning Study. *Scientific Studies of Reading, 22(3),* 191-208.

Ziegler, J., Castel, C., Pech-Georgel, C., George, F., Alario, F.-X., & Perry, C. (2008). Developmental dyslexia and the dual route model of reading: simulating individual differences and subtypes. *Cognition, 107(1),* 151–178.

Zorzi, M., Houghton, G. and Butterworth, B. 1998. Two routes or one in reading aloud? A connectionist dual-process model. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1131–1161.

Table 1 Fixed and random effects for performance in training task when comparing baseline, DF and CT conditions in pairs in Experiment 1

	Model 1		Model 2		Model 3a (baseline-DF)		Model 3b (baseline-CT)		Model 3c (DF-CT)	
Fixed Effects	EST	SE	EST	SE	EST	SE	EST	SE	EST	SE
Intercept	.50***	.05	.52***	.05	.52***	.05	.53***	.05	.53***	.05
Lin	.62***	.07	.66***	.08	.67***	.08	.67***	.08	.67***	.07
Quad	-.07	.04	-.06	.04	-.06	.04	-.06	.04	-.06	.04
MP			.40***	.07	.39***	.07	.40***	.07	.40***	.07
MP:Lin			.30**	.10	.30**	.10	.30**	.11	.30**	.10
MP:Quad			-.03	.08	-.04	.08	-.04	.08	-.03	.08
Base-DF					.01	.14				
Base-DF:Lin					.23	.19				
Base-DF:Quad					.16	.10				
Base-DF:MP					-.25*	.13				
Base-CT							-.42***	.12		
Base-CT:Lin							-.30	.18		
Base-CT:Quad							.13	.10		
Base-CT:MP							-.38**	.12		
DF-CT									-.44***	.12
DF-CT:Lin									-.52**	.18
DF-CT:Quad									-.00	.10
DF-CT:MP									-.15	.13
Random Effects	Var.		Var.		Var.		Var.		Var.	
PAR Intercept	.18		.12		.12		.10		.09	
Lin	.28		.23		.22		.21		.18	
Quad	.02		.01		.01		.01		.01	
PAR:MP Intercept			.15		.15		.14		.15	
Lin			.20		.20		.20		.19	
Quad			.04		.04		.04		.04	

Note. Baseline-DF = comparison between baseline and DF conditions; baseline-CT = comparison between baseline and CT conditions; DF-CT = comparison between DF and CT conditions; EST = parameter estimate; Lin = linear block term; Quad = quadratic block term; MP = mapping; CD = condition; PAR: participants; Var.: variance.
*‡ p < .10, * p < .05, ** p < .01, *** p < .001*

Table 2 Fixed and random effects for performance in Categorization Task when comparing baseline, DF and CT conditions in pairs in Experiment 1

		Model 1		Model 2		Model 3a (baseline-DF)		Model 3b (baseline-CT)		Model 3c (DF-CT)	
Fixed Effects		EST	SE	EST	SE	EST	SE	EST	SE	EST	SE
Intercept		.97***	.11	1.11***	.13	1.12***	.13	1.12***	.12	1.14***	.11
WT		.05	.06	.05	.06	.05	.06	.05	.06	.05	.06
MP				.56**	.18	.54**	.18	.55**	.17	.56**	.18
MP:WT				-.02	.12	-.02	.12	-.02	.12	-.02	.12
Base-DF						.38	.32				
Base-DF:WT						.19	.16				
Base-DF:MP						-.78‡	.45				
Base-CT								-.88**	.29		
Base-CT:WT								.01	.14		
Base-CT:MP								-1.23**	.40		
DF-CT										-1.29***	.27
DF-CT:WT										-.16	.15
DF-CT:MP										-.55	.44
Random Effects		Var.		Var.		Var.		Var.		Var.	
PAR	Intercept	.87		1.07		1.06		.93		.79	
	WT	.00		.01		.01		.01		.01	
	MP			2.01		1.92		1.69		1.89	

Note. Baseline-DF = comparison between baseline and DF conditions; baseline-CT = comparison between baseline and CT conditions; DF-CT = comparison between DF and CT conditions; EST = parameter estimate; WT = word type; MP = mapping; CD = condition; PAR: participants; Var.: variance.
*‡ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$*

Table 3 Fixed and random effects of block, mapping, reflective and reflexive learning for performance in training task in Experiment 2.

		Model 1		Model 2		Model 3a (RFLT)		Model 3b (RFLX)		Model 4 (RFLT & RFLX)	
Fixed Effects		EST	SE	EST	SE	EST	SE	EST	SE	EST	SE
Intercept		.63***	.05	.65***	.05	.65***	.05	.65***	.05	.67***	.05
Lin		.82***	.07	.87***	.08	.87***	.07	.87***	.07	.88***	.07
Quad		-.06	.04	-.04	.04	-.04	.04	-.04	.04	-.04	.04
MP				.39***	.07	.39***	.07	.39***	.07	.38***	.07
MP:Lin				.34**	.12	.34**	.12	.34**	.12	.34**	.12
MP:Quad				.02	.08	.02	.08	.02	.08	.02	.08
RFLT						.13*	.05			.11*	.05
RFLT:Lin						.14‡	.07			.13‡	.08
RFLT:Quad						-.05	.04			-.05	.05
RFLT:MP						-.00	.06			-.01	.06
RFLX								.14**	.05	.11*	.05
RFLX:Lin								.11	.07	.08	.07
RFLX:Quad								-.02	.04	-.01	.04
RFLX:MP								.00	.06	.00	.06
RFLT:RFLX										-.10‡	.05
RFLT:RFLX:MP										.02	.06
RFLT:RFLX:Lin										-.07	.08
Random Effects		Var.		Var.		Var.		Var.		Var.	
PAR	Intercept	.17		.11		.10		.09		.07	
	Lin	.24		.13		.11		.11		.10	
	Quad	.02		.01		.01		.00		.00	
PAR:MP	Intercept			.13		.13		.13		.13	
	Lin			.28		.28		.28		.28	
	Quad			.04		.04		.04		.04	

Note. RFLT = reflective learning; RFLX = reflexive learning; EST = parameter estimate; Lin = linear block term; Quad = quadratic block term; MP = mapping; PAR: participants; Var.: variance.
*‡ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$*

Table 4 Fixed and Random Effects of Block, Mapping, Reflective and Reflexive Learning for Performance in Categorization Task in Experiment 2

		Model 1		Model 2		Model 3a (RFLT)		Model 3b (RFLX)		Model 4 (RFLT & RFLX)	
Fixed Effects		EST	SE	EST	SE	EST	SE	EST	SE	EST	SE
Intercept		1.19***	.13	1.42***	.15	1.42***	.14	1.42***	.14	1.50***	.14
WT		.00	.07	-.02	.07	-.02	.07	-.02	.07	-.02	.07
MP				1.13***	.21	1.13***	.21	1.13***	.21	1.21***	.22
MP:WT				-.15	.13	-.15	.13	-.15	.13	-.15	.13
RFLT						.32*	.14			.29*	.14
RFLT:WT						-.01	.06			-.01	.07
RFLT:MP						.16	.21			.15	.21
RFLX								.23	.14	.14	.14
RFLX:WT								-.00	.06	-.00	.07
RFLX:MP								.11	.21	.05	.21
RFLT:RFLX										-.36*	.14
RFLT:RFLX:MP										-.34	.21
Random Effects		Var.		Var.		Var.		Var.		Var.	
PAR	Intercept	1.02		1.30		1.22		1.25		1.07	
	WT	.00		.01		.01		.01		.01	
	MP			2.40		2.36		2.40		2.29	

Note. RFLT = reflective learning; RFLX = reflexive learning; EST = parameter estimate; WT = word type; MP = mapping; PAR: participants; Var.: variance.
*‡ p < .10, * p < .05, ** p < .01, *** p < .001*

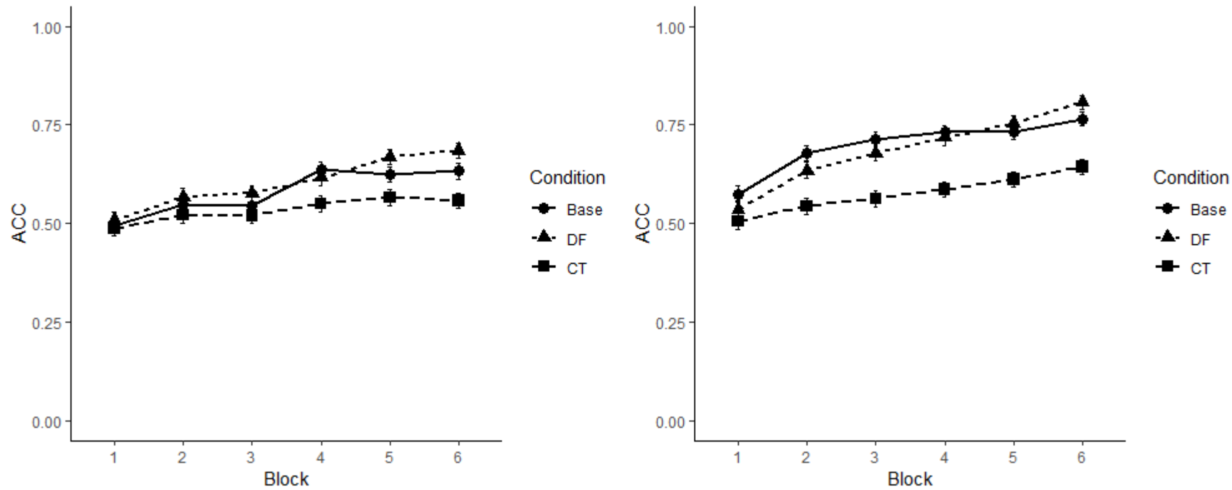

Figure 1. Training data across blocks in the three conditions in O-P (left panel) and O-S (right panel) mappings in Experiment 1.

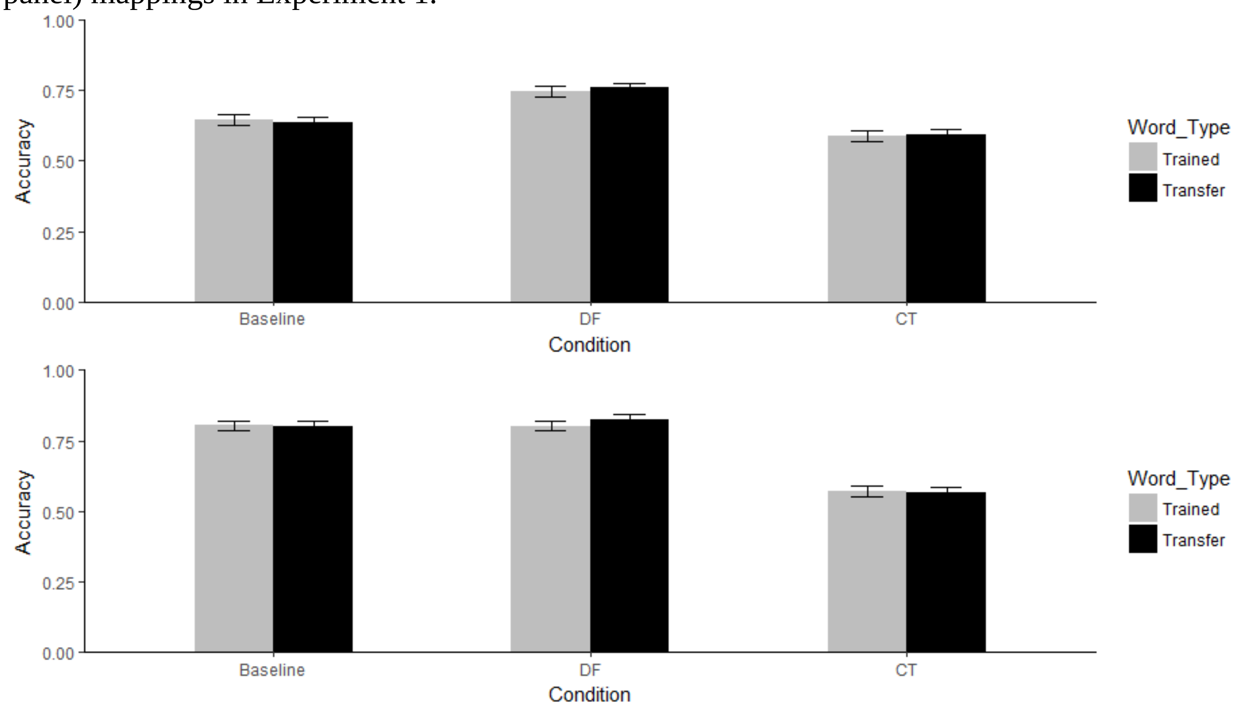

Figure 2. Performance in O-P (Upper Panel) and O-S (Lower Panel) mappings in baseline, DF and CT conditions in the Categorization Phase in Experiment 1

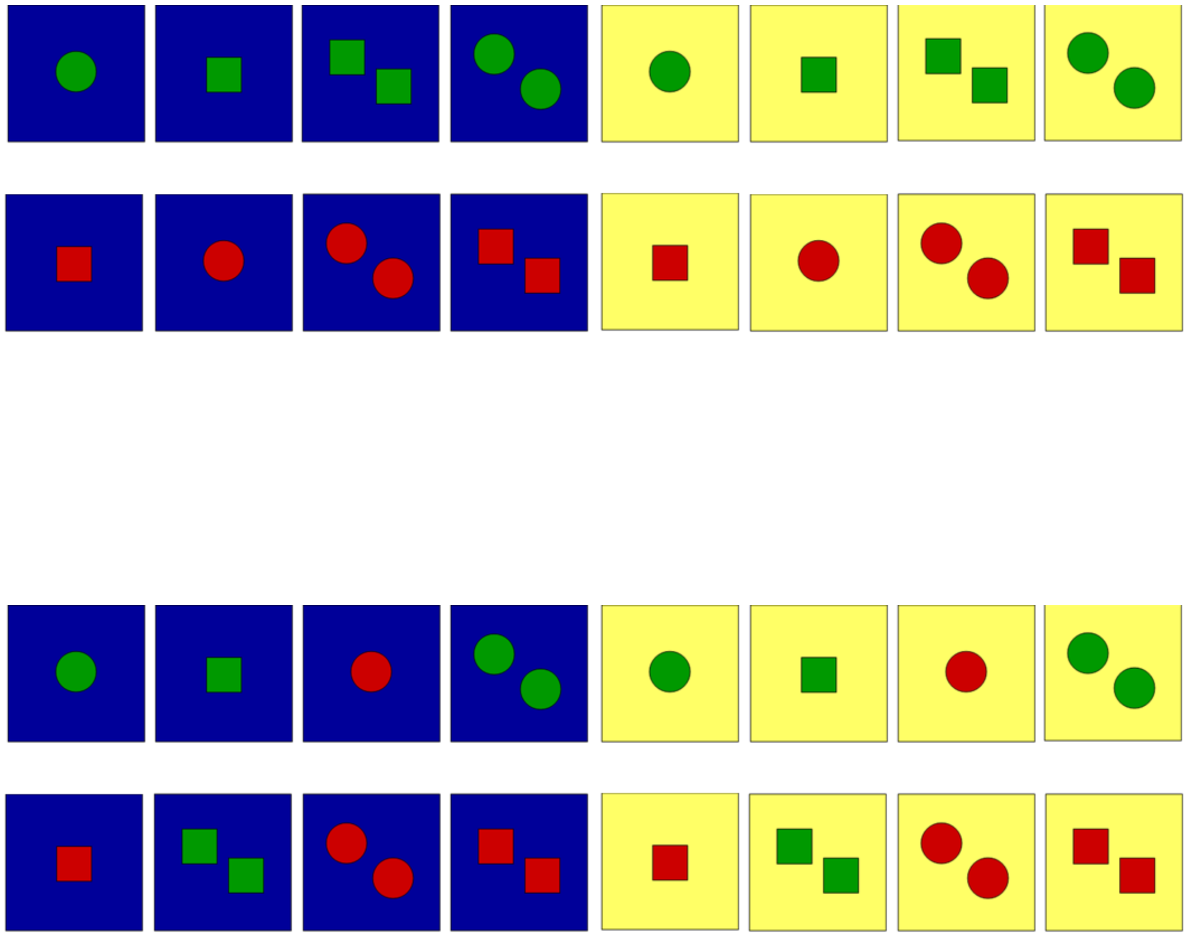

Figure 3. Examples of items used in reflective learning and reflexive learning.

Note: The upper group (upper two rows) is an example of reflective learning, with the items in the first row belonging to category A and items in the second row belonging to category B. Regularity in reflective learning was one-dimensional, and, in this example, the selected dimension was the symbol color in each item, i.e., if the symbol color was green, the item belonged to category A, otherwise category B. The lower group (lower two rows) is an example of reflexive learning, with the items in the first row belonging to category A and items in the second row belonging to category B. Regularity in reflexive learning was multi-dimensional: in this example, the selected dimensions were the shape, color, and number of the symbol(s) in each item, and the background color was irrelevant to the category. For shape, circle = -1, square = 1; for color, green = -1, red = 1; for number, one = -1, two = 1. If value (shape) + value (color) + value (number) was > 0, the item belonged to category A, otherwise category B.

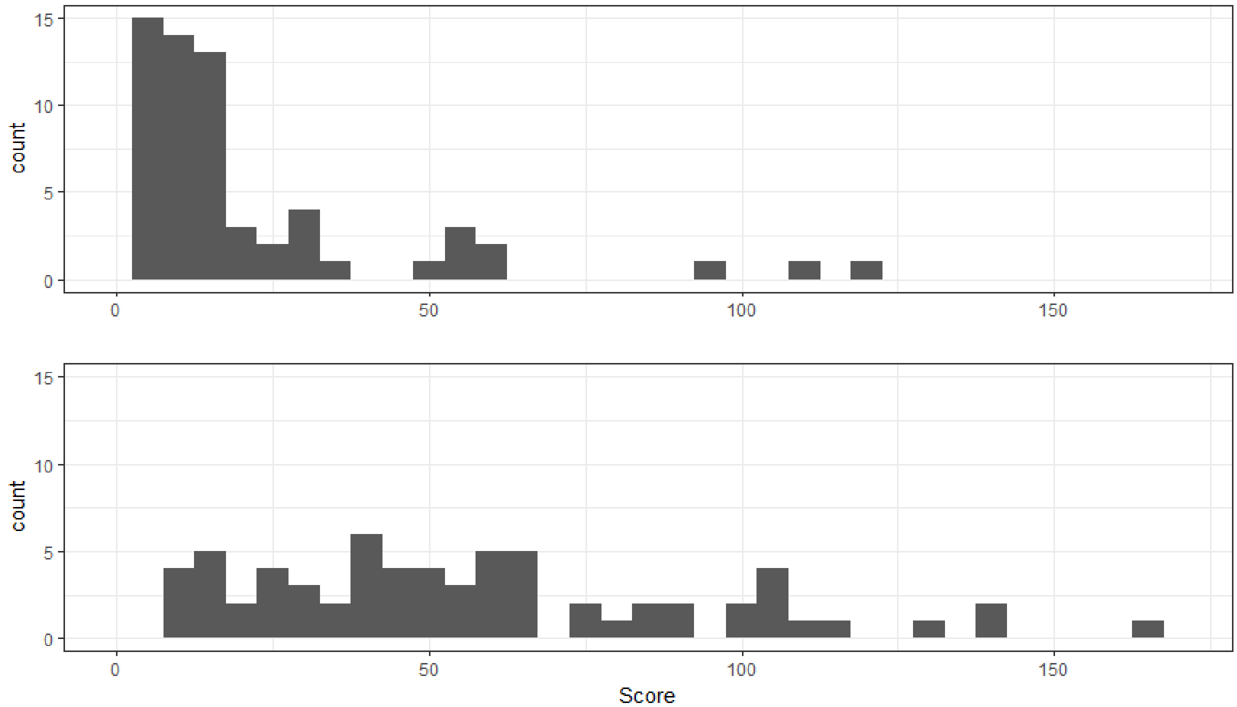

Figure 4. Distribution of reflective (upper panel) and reflexive (lower panel) visual category learning scores in Experiment 2.

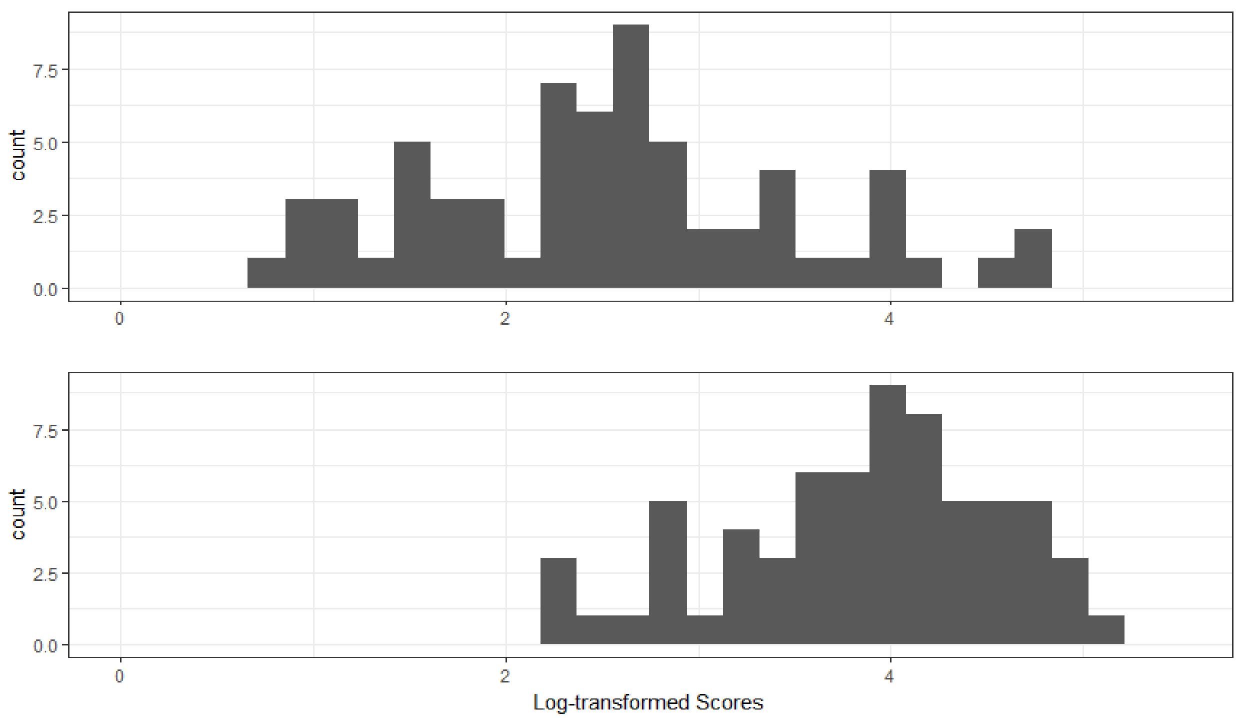

Figure 5. Distribution of log-transformed reflective (upper panel) and reflexive (lower panel) visual category learning scores in Experiment 2.

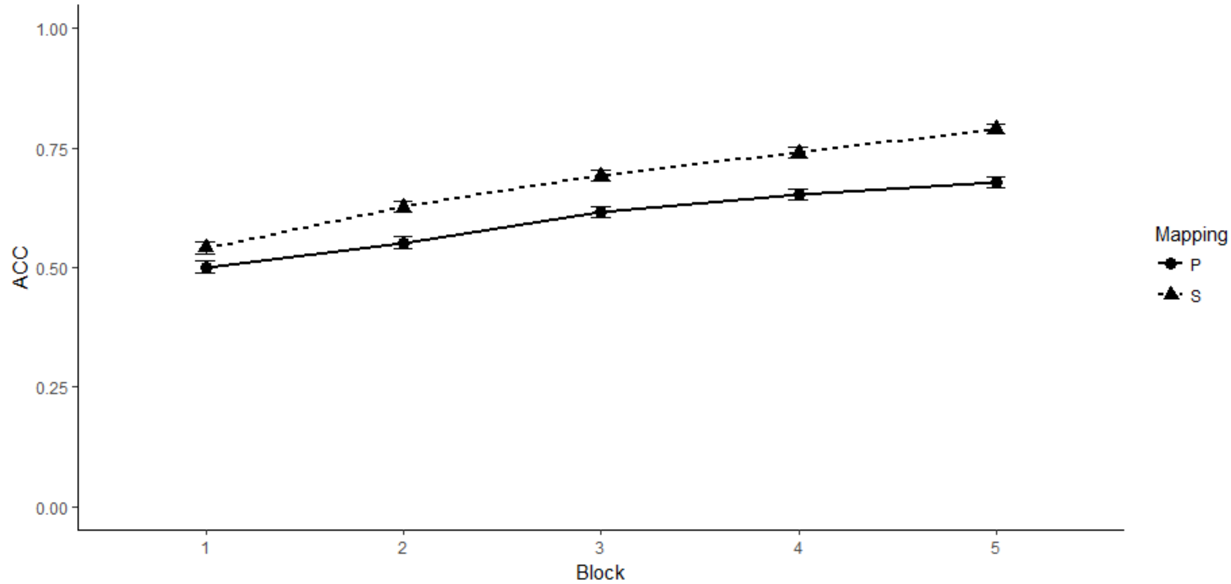

Figure 6. Training accuracy across blocks in O-P and O-S learning in the AOL task Experiment 2. (The legend should have O-P and O-S and not just P and S.)

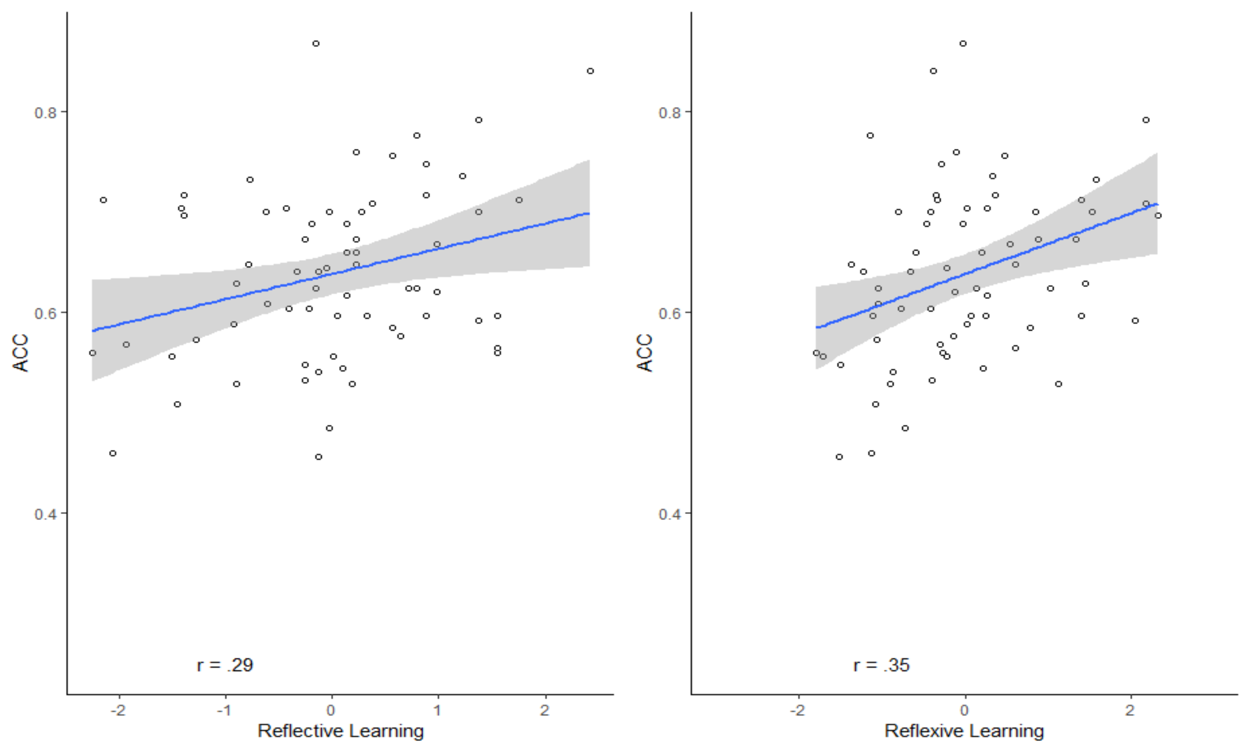

Figure 7. Scatterplot of training performance and reflective/reflexive learning (Note: Performance of training data was collapsed over blocks and mappings. Reflective/reflexive learning scores were log transformed, scaled and multiplied by -1.)

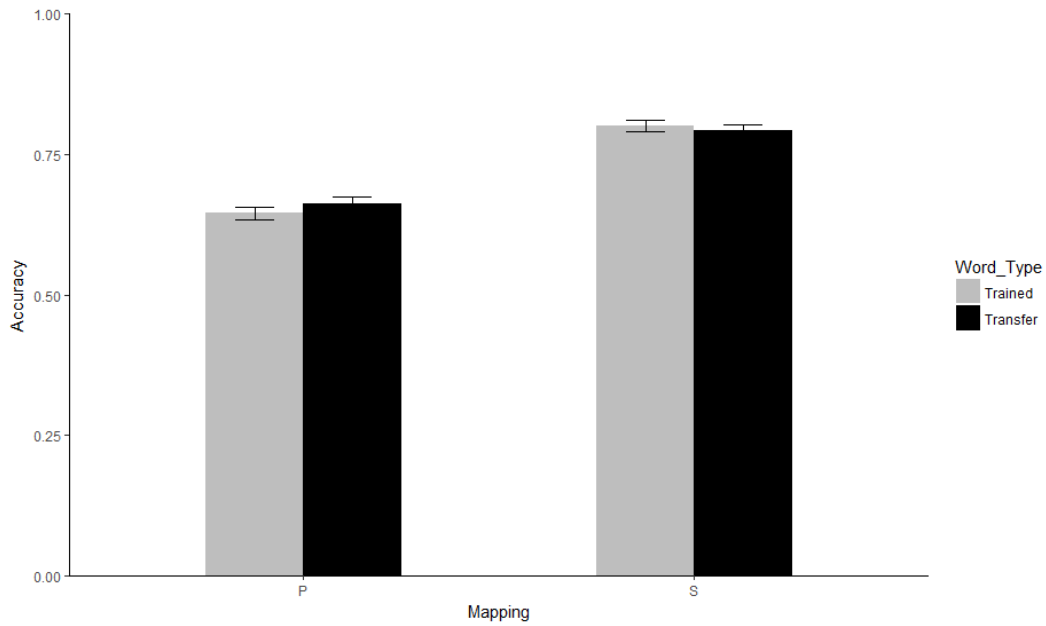

Figure 8. O-P and O-S learning in the categorization tasks in Experiment 2. (bars should be labeled O-P and O-S, not P and S)

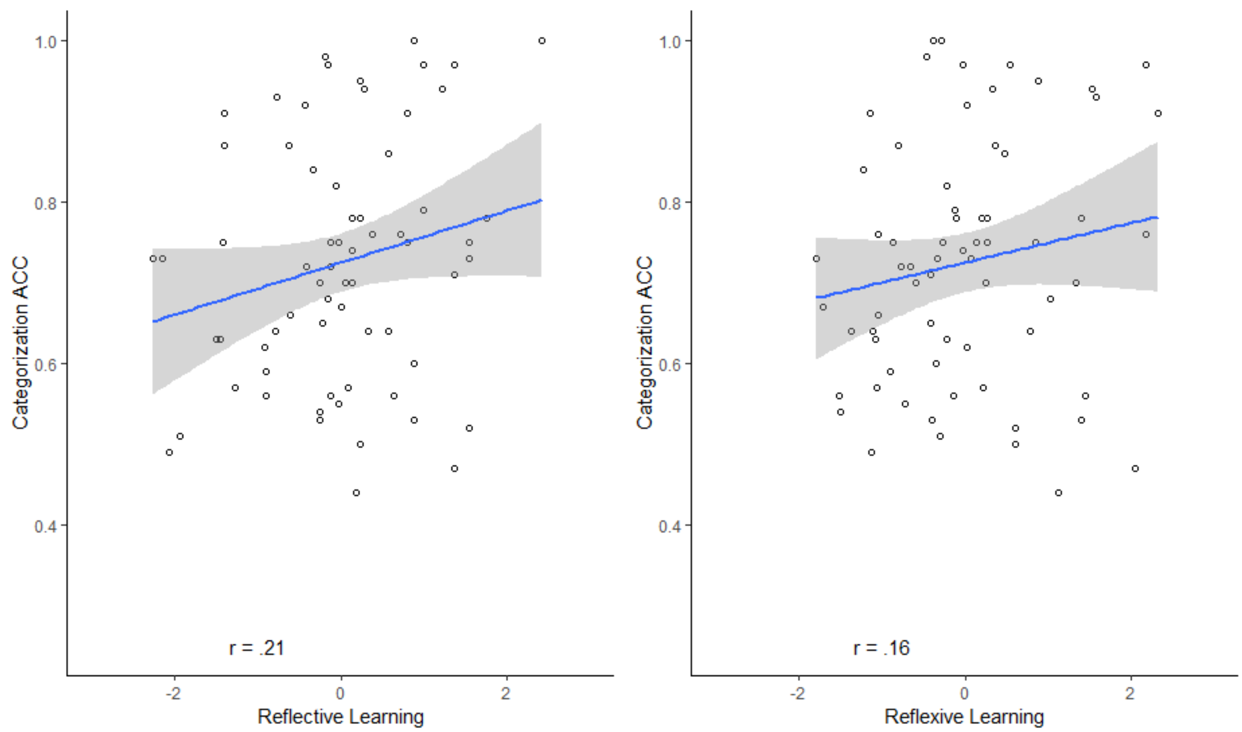

Figure 9. Scatterplot of categorization performance and reflective/reflexive learning (Note: Performance of categorization data was collapsed over word type. Reflective/reflexive learning scores were log transformed, scaled and multiplied by -1.)

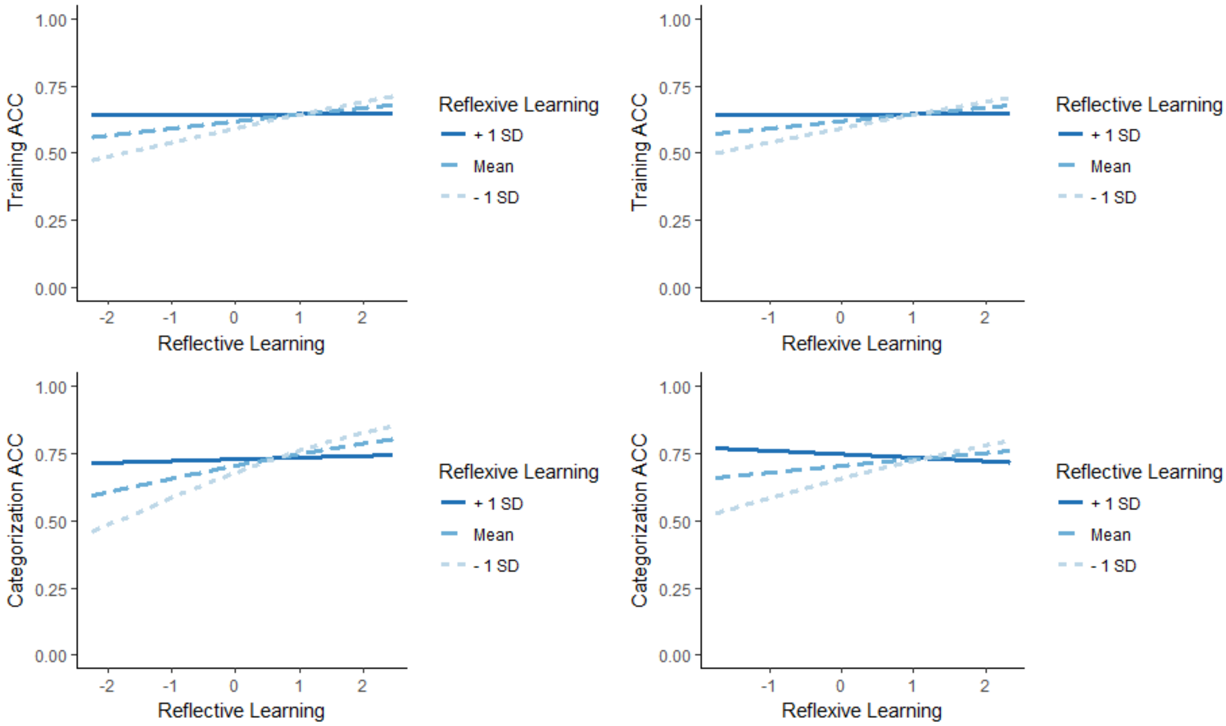

Figure 10. Interactions between reflective and reflexive learning for training task (upper panels) and for categorization (lower panels) in Experiment 2.

Appendix 1. Semantic and Phonological Radicals Used in Experiment 1 and 2 Training Task

Pho \ Sem	士 (animal)	中 (furniture)	下 (fruit)	刀 (clothing)	山 (body part)
小	dog	bed	pear	shoe	nose
大	cat	desk	banana	coat	leg
儿	horse	lamp	peach	shirt	arm
广	cow	chair	grapes	sock	eye
父	lion	table	apple	pants	foot

Semantic radicals were in the first row and phonological radical in the first column. Each semantic radical indicated a semantic category, as shown in the first row. The English word in each cell indicated the meaning of the word composed of the two radicals in the corresponding column and row, e.g., the meaning of 尗 and 尘 was “dog”.

Pho \ Sem	士	中	下	刀	山
小 /-eɪs/	/bleɪs/	/deɪs/	/weɪs/	/neɪs/	/teɪs/
大 /-æd/	/flæd/	/spæd/	/stæd/	/væd/	/træd/
儿 /-ɜrb/	/dɜrb/	/nɜrb/	/tɜrb/	/wɜrb/	/mɜrb/
广 /-aɪv/	/spaɪv/	/zaɪv/	/paɪv/	/maɪv/	/braɪv/
父 /-ʌk/	/nʌk/	/wʌk/	/brʌk/	/prʌk/	/drʌk/

Semantic radicals were in the first row and phonological radical in the first column. Each phonological radical indicated a specific rhyme, as shown in the first column. The phonetic symbol in each cell indicated the pronunciation of the word composed of the two radicals in the corresponding column and row, e.g., the pronunciation of 尗 and 尘 was “/bleɪs/”.